

nature



GENES VS. ADRENALINE
A few fibres in the
amygdala
ARE A MESSY
Building on the high
ground in depression
GENETIC DISEASES
A no-nonsense drug

THE IDLING BRAIN

Neural circuits revealed by
spontaneous fluctuations

NATURE JOURNALS
www.nature.com/nature

South Africa's AIDS plan

A strategic plan has been developed to tackle the nation's HIV crisis — at last.

After years of prevarication, the government of South Africa has finally adopted a sound and comprehensive strategy for managing the country's AIDS epidemic. The 160-page plan, which was endorsed by the South African National AIDS Council on 30 April, lays out proposals to cut infection rates, improve diagnosis of the disease, and treat the estimated 5.5 million South Africans already infected with HIV.

The plan has been welcomed by physicians, AIDS researchers and activists in South Africa, who have long been frustrated by the reluctance of President Thabo Mbeki's government to confront the world's worst national AIDS problem head-on.

It sets two main objectives: by the end of 2011, the rate of infection should be halved; and antiretroviral treatment should be readily accessible to 80% of the people who require it. The estimated cost of all this is US\$6 billion over the next five years, with drug costs accounting for 40% of the total. The plan expresses the hope that the private sector and foreign donors will provide half of the cash, with the South African government paying the rest.

According to a report released in November by the Actuarial Society of South Africa (ASSA), 1.8 million people have died of AIDS in South Africa so far. Current estimates of the number of people requiring antiretroviral treatment range between 800,000 and 1 million. But only about 300,000 are receiving it, with two-thirds of them being treated by the Department of Health and the rest by private healthcare schemes and non-governmental organizations. Many more of those already infected with HIV will need drug treatment within five years.

The absence of an effective AIDS strategy for South Africa until now has particularly affected the availability of treatment for children born with the disease. Only one-tenth of the estimated 200,000 children who might benefit from antiretroviral treatment are receiving it, according to the International Treatment Preparedness Coalition. That is partly because the treatment of children carries extra bureaucracy. Just to be tested for HIV (a prerequisite for access to antiretroviral treatment), they require identity numbers, which

in many cases they don't have, as well as consent from a biological parent — a problem because many are cared for by other members of the extended family. There is no other medical condition where such criteria have to be met before starting treatment.

Slowing the infection rate may be even more challenging than administering therapy. The rate is already falling in urban areas: ASSA estimates that 530,000 people were infected in South Africa last year, down from the peak annual infection of 650,000 in 1998. But in rural areas such as the Umkhanyakude area of KwaZulu-Natal province, for example (see page 26), infection rates remain high and public understanding of AIDS is threadbare.

Additionally, a lack of trained doctors and nurses is likely to act as a serious constraint on the full implementation of the strategic plan. Many have grown frustrated over the government's AIDS policies, sometimes to the point of leaving the country, or at least the public sector.

Although admirably broad in its scope, the plan falls short of addressing some of the technical and bureaucratic obstacles that stand in the way of AIDS prevention and treatment. Another problem is Mbeki's failure, so far, to publicly acknowledge the AIDS crisis.

Indeed, reluctance on the part of some of South Africa's leaders to face up to the extent of the problem remains an issue. But deputy president Phumzile Mlambo-Ngcuka, whom many see as a possible successor to Mbeki when his term expires in 2009, has rewritten the government's AIDS agenda by helping to draw up the strategic plan. The health minister Manto Tshabalala-Msimang, who embarrassed South Africa with her remarks at the World AIDS Congress in Toronto last August (see *Nature* 444: 663; 2006), is currently on sick leave. If Mbeki is serious about confronting AIDS, he should appoint someone with the necessary aptitude and enthusiasm, such as deputy health minister Nozizwe Madlala-Routledge, to succeed her. ■

"A lack of trained doctors and nurses is likely to constrain the full implementation of the plan."

Share your lab notes

The use of electronic laboratory notebooks should be supported by all concerned.

Too often when errors or cases of fraud occur in science, the lab data required to reconstruct what happened have gone astray. And too often, the co-authors failed to exert due scrutiny on their colleagues' activities in order to prevent such misfortunes. The damage to personal and institutional reputations can be severe and, in rare high-profile cases, public trust can be eroded.

It is therefore in everyone's interest to pre-empt such cases as far as possible. Electronic laboratory notebooks offer a partial solution — and have other advantages too. This is despite the fact that maximizing their benefits will require a change in culture that many researchers will no doubt initially resist.

Electronic notebooks, like their paper cousins, record the daily thoughts and experiments of bench scientists. Ideally, they contain data that flow automatically from lab instruments and can be read by all lab members. Pages are date- and time-stamped, and all changes tracked and signed. Earlier versions can be reconstructed.

There are numerous e-notebook products available, but none dominates in all sectors. The pharmaceutical industry, which is well

accustomed to regulation, has adopted company-wide solutions, and the US Food and Drug Administration has determined that the use of electronic notebooks is acceptable in drug filings. This high degree of usage is in stark contrast to academia.

So why bother? Most importantly, e-notebooks allow the sharing of data, to the immediate benefit of collaborators (for examples, see *Nature* **436**, 20; 2005). And just knowing that a notebook is available to others in the lab, and archived for the future, should compel the keeping of better records.

But one can and should go further. Electronic notebooks can be archived by researchers' employers, with a number of attendant benefits. If each notebook (or subset of it) is allocated a unique identifying code — a permanent alphanumeric string containing information about provenance, creation dates and digital location — it can be cited in journals as a confirmation that the data are safely stored, ultimately available and sharable (with due regard for the rights of the researchers involved). It also confirms that the original data can be retrieved in the case of errors or accusations of fraud. No longer would claims of lost notebooks be brought up in misconduct investigations.

An additional benefit is that the data may have a value not only to

the researchers who produced them but to others too, independently of the publications that report them. That value can be recognized explicitly by citation of the identifying code, enabling due credit to be given to the researchers who produced them.

Any change that requires adjustments to the way scientists go about their daily work meets with resistance and, for many researchers, lab notebooks are wrongly considered to be private property. Academic acceptance of e-notebooks will not improve unless universities promote their use and recognize that e-notebooks can help them fulfil their responsibilities as the owners of most grant-funded data. There will be challenges with respect to maintenance of the archives and the standardization of software, but those who have invested in this approach have seen the benefits.

Institutions therefore need to show leadership in this area, and funding agencies should provide additional infrastructure support earmarked for the development and upkeep of electronic notebook systems. Funding agencies also need to recognize that, by providing such support, some of the concerns over the loss of data can be assuaged, and the rigour and transparency of publicly funded research will be improved. ■

Hard to defend

US missile defence plans require scrutiny.

The utility and value of missile defence systems remain unproven, but the United States plans to go ahead and deploy one in Eastern Europe regardless. The plan has delighted some former satellites of the Soviet Union, such as Poland, but irked other European nations, and remains of questionable relevance to US defence needs. Unfortunately, no one in Congress is saying so.

Defence officials say that ten interceptors in Poland and a radar in the Czech Republic are needed to protect the United States and its allies from yet-to-be developed Iranian ballistic missiles. But Russia's President Vladimir Putin believes that the shield has other purposes. On 27 April, shortly after a visit from US defence secretary Robert Gates, Putin announced that he would retaliate against the missile defence plan by pulling out of the 1990 Treaty on Conventional Armed Forces in Europe. US allies in Western Europe are irritated, meanwhile, that negotiations over the expansion will take place through bilateral channels, circumventing NATO.

But Democrats in Congress, who have long been suspicious of the missile defence system, are reluctant to oppose it. At Senate hearings last week, the Missile Defense Agency asked for nearly \$9 billion in the next financial year, most of it to purchase early-warning radars and long-range interceptors, including some of the equipment for the European sites.

The panel's chairman, Daniel Inouye (Democrat, Hawaii), told the agency's leaders that they should be "proud" of their achievements and called for a beefing up of the system's capability. Two of the system's opponents, Dianne Feinstein (Democrat, California) and Byron Dorgan (Democrat, North Dakota), did raise specific concerns about its ability to detect decoys. But nobody was ready to side

with Putin and publicly question the agency's plans to field the new interceptors.

That's unfortunate, because the missile defence project remains as technically questionable and strategically undesirable as ever. In tests, the interceptors have had mixed success at striking incoming warheads. Physicists have long been cautious about their effectiveness, and advocacy groups such as the Union of Concerned Scientists and the Federation of American Scientists continue to express deep scepticism over whether such a system can ever stop an incoming ballistic-missile attack. They think the system remains vulnerable to decoys and other relatively simple countermeasures.

Furthermore, this brittle defence can do nothing to stop cruise missiles of the sort already developed by Iran, nor can it intercept a warhead smuggled into the country on a truck or boat. In today's security environment, these threats are more direct than the long-range missile capability that Iran and North Korea do not yet have.

But last week's hearing suggests that missile defence may have become a pariah issue for the Democrats. Like gun control — an idea that cannot speak its name, even in the wake of last month's horrific shooting at Virginia Tech — the abandonment of missile defence seems to be politically out-of-bounds. Most leading Democrats are reluctant to oppose the programme — a form of defence that some polls suggest Americans mistakenly think they already have — in case it makes them look soft on national security. Nobody wins elections by dispelling illusions of security, and a few billion dollars is a small price to pay for political expediency, at least where Pentagon budgets are concerned.

But if the system is sham, as well as a menace to foreign relations, no more money should be wasted on it. Instead of humbly waving the programme's massive budget through, Congress should be asking the Government Accountability Office to establish exactly what US taxpayers have got for the tens of billions of dollars that the Bush administration has expended in pursuit of this myopic vision. ■

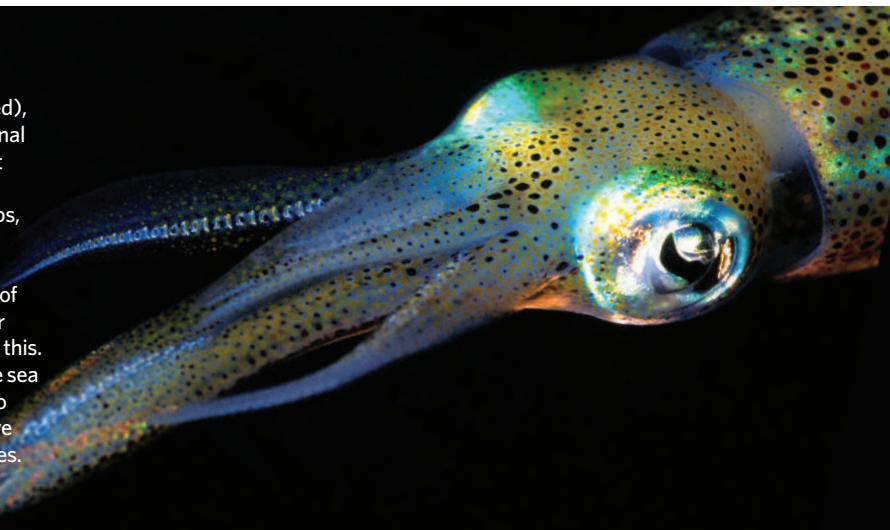
RESEARCH HIGHLIGHTS

Squids in

Nature Struct. Mol. Biol. **14**, 427–431 (2007)

Some marine animals, such as the squid (pictured), cope with life in the ocean by allowing their internal fluids to be as salty as the surrounding water. But high extracellular sodium concentrations could interfere with the cells' sodium-potassium pumps, which drive the import of some nutrients.

Miguel Holmgren of the National Institutes of Health in Bethesda, Maryland, Joshua Rosenthal of the University of Puerto Rico in San Juan and their colleagues have discovered how the squid avoids this. Having deciphered the molecular structure of the sea creature's pumps, the team found that four amino acids around the external mouth of the pump have a more strongly positive charge than in vertebrates. This prevents backflow of sodium into the cell by repelling the positively charged sodium ions.



BRANDON COLE/MARINE PHOTOGRAPHY/ALAMY

NEUROSCIENCE

Light-headed

Nature Neurosci. doi:10.1038/nn1891 (2007) and *Neuron* **54**, 205–218 (2007)

A light-activated ion channel found in algae, known as channelrhodopsin-2, has enabled two teams to map neural circuits in the brains of mice. The researchers transferred the ion channel into discrete sets of neurons, which then fired in response to light.

By simultaneously stimulating and recording from the neurons, Karel Svoboda at the Howard Hughes Medical Institute in Ashburn, Virginia, and his colleagues mapped part of a circuit in the somatosensory cortex, which processes information about touch.

Michael Ehlers at Duke University Medical Center in Durham, North Carolina, and his colleagues studied a part of the brain involved in conveying information about odours to the cortex. To do this, Ehlers's team created transgenic mice that express the ion channel, whose offspring can be used in future rounds of experiments.

ORGANIC CHEMISTRY

The colour purple

Chem. Commun. doi:10.1039/b618926a (2007)

The compound that launched the modern chemical industry is more complex than anyone had realized, say scientists in Portugal.

João Seixas de Melo of the University of Coimbra and his co-workers have reproduced the synthesis of 'mauveine' conducted in 1856 by William Perkin. The purple substance that Perkin produced was the first of the aniline dyes, on which many of

today's big chemical companies were founded.

It was discovered in 1994 that the dye contains two molecules, dubbed mauveines A and B. Now de Melo and his colleagues report that their synthesis yields two further variants of the basic mauveine structure, also present in historical dye samples. Perkin's glorious mauve is therefore a subtle blend of the molecules' slightly different colours.

NANOTECHNOLOGY

New balls, please

Phys. Rev. Lett. **98**, 166804 (2007)

Carbon 'buckyballs' are facing competition from boron balls: new calculations predict that boron can assemble into spherical molecules, too.

Buckyballs are cage-shaped molecules made from 60 carbon atoms arranged in a pattern of hexagons. Their round, hollow

structure is enticing to researchers, who have turned the molecules into wheels for molecular cars, as well as exploring more serious applications in composite materials and electronics. The most stable of the predicted boron balls — which should have similar technological appeal — contains 80 atoms (pictured below).

Boris Yakobson of Rice University in Houston, Texas, and his colleagues say that their calculations can't predict how easy this molecule might be to make, but encourage experimenters to try.

RNA BIOLOGY

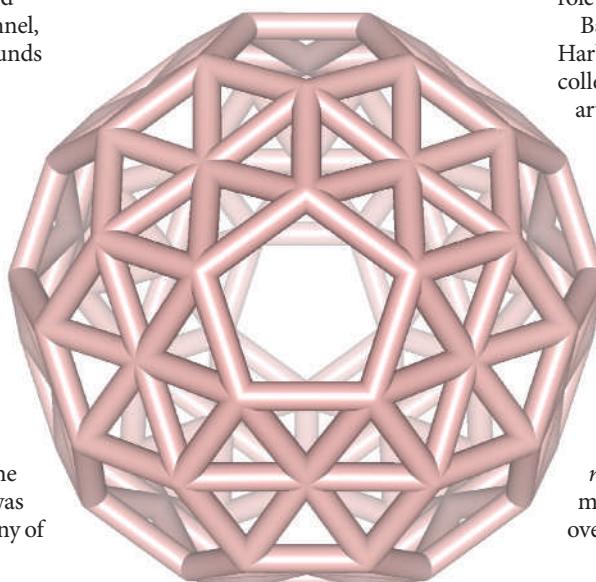
Tiny heartbreakers

Nature Med. **13**, 486–491 and 613–618 (2007)

Tiny RNA molecules known as microRNAs regulate everything from development to stress responses — they also play an important role in heart function, two studies report.

Baofeng Yang and Zhiguo Wang of the Harbin Medical University in China and their colleagues report that people with coronary artery disease express more of a microRNA named *miR-1*. Overexpressing *miR-1* in rats promoted irregular heartbeats; inhibiting *miR-1* relieved the condition.

Meanwhile, a team led by Cesare Peschle of the Italian National Institute of Health in Rome and Gianluigi Condorelli of the University of California, San Diego, has shown that patients with a pathological thickening of heart muscle known as cardiac hypertrophy produce less of another miRNA, *miR-133*. Inhibiting *miR-133* in mice triggered the condition, whereas overexpressing *miR-133* prevented it.



O. PUPYSHEVA/RICE UNIV.

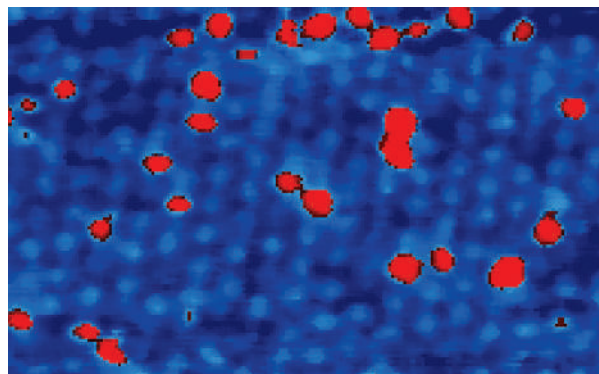
MICROBIOLOGY

Respiration in reverse

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0610456104 (2007)

Microbial geneticists have sequenced the genome of a bacterium that plays a key role in the global anaerobic carbon cycle, in a bid to study its unorthodox lifestyle. *Syntrophus aciditrophicus* runs its electron-transport chain — a crucial set of respiratory reactions — backwards.

Researchers led by Robert Gunsalus of the University of California, Los Angeles, sequenced *S. aciditrophicus*'s 3.2 million letters of DNA code. They discovered genes that underpin the bacterium's consumption of organic molecules such as alcohols and fatty acids, and that allow its electron transport to function in this unusual direction.



CHEMISTRY

Work in progress

Nature Nanotechnol. doi:10.1038/nnano.2007.106 (2007)

Snapshots of catalytic molecules at work have been taken with a scanning tunnelling microscope. Johannes Elemans of Radboud University Nijmegen, the Netherlands, and his team have watched individual disk-shaped organic molecules known as porphyrins as they add oxygen atoms to alkenes.

A manganese ion at the centre of each porphyrin seizes an oxygen atom from molecular oxygen (O_2) before inserting it into the alkene. The image (left) reveals which of the porphyrins on a gold surface have an oxygen, showing that these tend to come in pairs: the two oxygens from an O_2 molecule are captured by adjacent porphyrins.

QUANTUM INFORMATION

Which way did it go?

Phys. Rev. Lett. **98**, 160403 (2007)

Even by the surreal standards of quantum physics, the result predicted by Lev Vaidman of Tel-Aviv University in Israel is bizarre: a quantum particle can arrive somewhere without having taken the route that leads there.

Vaidman reached this conclusion after examining a recent proposal for 'counterfactual computation' (O. Hosten *et al.* *Nature* **439**, 949–952; 2006), in which a quantum computer performs a computation without running the algorithm it involves — where 'running the algorithm' means passing a photon through the computer.

In the set-up proposed by Hosten *et al.*, the photon is directed away from the device, but, nevertheless, measurement of the state of the photon at the end of the procedure yields the

result of the computation. Vaidman shows that a photon sent away from the device can still reach and pass through it. Thus, the computer does actually run the algorithm and the computation isn't 'counterfactual'.

CANCER

Aid to resistance

Science doi:10.1126/science.1141478 (2007)

The drug gefitinib is at first an effective treatment for certain types of lung cancer, but most cancers become resistant to it. New research provides insight into one mechanism by which this happens.

Gefitinib inhibits an enzyme known as EGFR kinase, and around half of the resistance cases are due to mutations in this target. Among the rest, Pasi Jänne of the Dana-Farber Cancer Institute in Boston, Massachusetts, and his colleagues found that some resistant cells showed increased expression of a tyrosine-kinase receptor named MET. This activates the same signalling pathway that EGFR kinase triggers — a pathway that mediates cell survival.

The researchers found that the *MET* gene had multiplied rather than mutated, and that inhibiting *MET* restored the cells' sensitivity to gefitinib.

CELL BIOLOGY

Algal aptitude

Genes Dev. **21**, 1190–1203 (2007)

MicroRNAs have shown up for the first time in a unicellular organism, suggesting that they evolved before multicellularity.

Previous searches in yeast and bacteria have failed to find any microRNAs. But Yijun Qi of the National Institute of Biological Sciences, Xiu-Jie Wang of the Chinese Academy of Sciences, both in Beijing, and their colleagues have now isolated some from the unicellular green alga *Chlamydomonas reinhardtii*.

The team showed that several of these microRNAs cleave specific messenger RNAs, a key part of the mechanism by which microRNAs regulate gene expression. What's more, expression of some microRNAs changed during gamete formation, suggesting a role in the reproductive process.

JOURNAL CLUB

Phil Bland
Imperial College London, UK

A planetary scientist learns how comet dust gets from the inner to the outer Solar System.

I was lucky enough to be part of a team studying the grains of comet dust collected last year by NASA's Stardust mission. Comets are primitive, pristine objects, and the Stardust samples are changing the way we think about how our

Solar System formed.

Among many surprising findings, perhaps the most significant is that a large fraction of the dust grains are minerals formed at high temperatures — temperatures expected only in the inner Solar System. How did this stuff get out to where the comet began its life, in the cold, outer regions of the Solar System?

At the recent Lunar and Planetary Science Conference in Houston, Texas, I learned about a numerical simulation that

potentially offers a neat solution (F. J. Ciesla and J. N. Cuzzi, abstract at www.lpi.usra.edu/meetings/lpsc2007/pdf/1386.pdf).

Observations of dusty disks around young stars show an inward flow towards the central star. Ciesla and Cuzzi's simulation suggests that this inward transport is confined to the top and bottom of the disk. It predicts that there is a narrow region near the disk's midplane where dust flows outwards — a flow sufficient to account for the Stardust results.

So now we know that comets contain a mixture of stuff from the inner Solar System, and we have a physical model that can explain how it got there. But we're still left with one question.

Virtually everything in the inner Solar System — Earth, Mars, the Moon, almost all meteorites — is depleted in volatile elements, which can't condense at high temperatures. But the cometary dust grains don't show this depletion signature. Why not? It'll be fun finding out.

NEWS

The race to wire up the poor

Nicholas Negroponte's much hyped 'One Laptop Per Child' (OLPC) scheme inspires or infuriates according to taste. The idea of revolutionizing education by distributing millions of '\$100 laptops' to children in the world's poorest countries can be seen as Soviet-style social engineering or as visionary California dreaming, mass empowerment or pointless frivolity. However you see it, though, the project is facing real problems — and competition.

Negroponte, who founded the MIT Media Lab in Cambridge, Massachusetts, and heads the non-profit OLPC Foundation, admitted last week that in the near term the laptops will cost not US\$100, but at least \$175. And although he is still confident that the first 3 million machines will be shipped by the end of 2007, an earlier estimate for the same timeframe was 10 million.

Meanwhile, as PC markets in industrialized nations become saturated, Intel, Microsoft, and other computer companies are turning their sights on developing countries, hoping to find the next billion computer users.

"We are at the most critical stage of OLPC's life," Negroponte told a meeting with industry analysts in Cambridge, Massachusetts, on 26 April. "A year and a half ago, we were selling a dream, but it's easy to sell dreams if you're passionate and can share that passion with other people. But that was dreams, and now we have got to launch."

Unique approach

To launch, Negroponte needs money. OLPC has an unorthodox business model: it sells only to governments and demands minimum orders of 250,000 machines at a time — down from its original demand of one million units. The idea is to achieve scale instantaneously, and to pass distribution and marketing costs on to the purchaser.

How much those extra costs will push the bill up is not clear. The OLPC project is based on the idea that just providing computers can make a huge difference, as children will spontaneously use them to make up for what they lack in books and other equipment and experiences. But many education researchers dispute this theory. And the more that classroom structure and teacher training is needed to maximize the computers' benefit, the more the programme will end up costing governments.

So far no countries have actually paid for their computers, let alone revamped their education efforts to prepare. Libya has come closest, signing a memorandum of under-



Nicholas Negroponte hopes to get computers to children in developing countries through his One Laptop Per Child scheme.

standing to buy 1.2 million laptops. Rwanda and Uruguay have said that they are committed to the project, and several other countries have expressed interest. But some are firmly ruling themselves out. "We cannot visualize a situation for decades when we can go beyond the pilot stage," Sudeep Banerjee, India's education secretary, said last year. "We need classrooms and teachers more urgently than fancy tools."

In March, in an investment report on OLPC and related companies — such as Quanta, the Taiwanese firm contracted to build the foundation's \$100 laptop — investment analysts at Merrill Lynch concluded that "despite the hype, we urge investors not to chase stocks related to OLPC". Low-end laptops can already be made at a cost of \$400, the report pointed out, and OLPC's lower costs come from using less-powerful components and eliminating

distribution and after-sales service, both of which impose burdens on the end user.

The report also points out that OLPC is not the only show in town. Intel is rolling out its \$400 Classmate PC to schools in Brazil and Mexico this year, and has pilot schemes in 25 other countries. Over the next 5 years it

is investing \$1 billion to promote access to information technology in developing countries. Microsoft, loathe to see the next billion computer users running the open-source Linux operating system that the OLPC computers use, last week announced that it

would offer developing countries a software package that includes XP and Office for \$3 for every new school PC bought.

No one denies that the rugged OLPC laptop contains some serious technological innovation. It runs on just 2 watts of electric-

"We need classrooms and teachers more urgently than fancy tools."

M. EULER/AP



QUANTUM CRYPTOGRAPHY IS HACKED

It is possible to eavesdrop on encrypted messages.
www.nature.com/news

ity — around a twentieth of a typical laptop's needs — and can be charged off-grid in various ways. Its display can be read in bright sunlight, and nearby laptops automatically create a wireless network between themselves with which, among other things, they can share an Internet connection.

Ashok Jhunjhunwala, a former director of the Indian Institute of Technology in Chennai, and one of the India's top experts in information technology sees the OLPC machine as "a great technological effort". But "proper support is a must. Nothing works without support, management and maintenance".

Change in focus

Poor countries have other public spending priorities, Jhunjhunwala argues, and initiatives such as OLPC should start where they can have a real effect — among the market of emerging middle classes and schools that can afford the computers for themselves. The OLPC will otherwise be irrelevant to developing countries; in the poorest strata of society "this toy will just be sold or stolen".

Jhunjhunwala himself is the brains behind the netPC and netTV computers built by Novatium, a Chennai-based company, which embraces the idea that PCs no longer need large, expensive hard drives and stacks of memory, but can instead act as gateways to computing power elsewhere on the Internet. Users are expected to access free web-based word-processing and spreadsheet software and storage, as offered by web services such as Google Office, or to take out a subscription for Microsoft Office and other paying software served over the web by Novatium's own central servers.

Whatever the future for the OLPC project, the fact that the next billion computer users will be in developing countries looks set to drive a new phase in computing innovation — one that could have repercussions in developed markets too. Last week, for the first time, Negroponte said that he was considering selling the computers to US children via their state governments, an idea he had previously rejected. The educational information appliance of the future might turn out to be a mobile telephone, or some sort of hybrid, but no one can rule out Negroponte's vision of one laptop per child coming true worldwide. That does not mean, though, that his specific approach will get much further. ■

Declan Butler

Long odds on a long shadow

Astronomers may make a major step forward in studying planets around other stars as early as next week — but the odds are strongly against them.

Last week's announcement by Stéphane Udry and his colleagues at the Geneva Observatory that they had found a planet in the 'habitable zone' of the red dwarf Gliese 581 caused a great deal of excitement (S. Udry *et al. Astron. Astrophys.*; in the press). Its low mass (about five times that of Earth) and its position in the habitable zone, where temperatures are compatible with water being present as a liquid, mean that Gliese 581 c, as the planet is called, could be more Earth-like than anything else yet seen beyond the Solar System.

However, finding out whether Gliese 581 c is a larger version of Earth or a smaller version of Neptune — a much less homely prospect — means measuring its density. The only way this can be done is by observing the planet cross the face of Gliese 581 and thus working out its radius. Unfortunately,

such a transit will be visible from our Solar System only if we happen to be sitting in the plane of the planet's orbit. The chances of that being the case are a daunting 50 to 1 against.

Nevertheless, any odds are worth taking when it's the only game in town. On 26 April, just three days after the planet's existence was revealed, Dimitar Sasselov of Harvard University began using the small Canadian space telescope MOST (Microvariability and Oscillations of Stars) to observe Gliese 581. Sasselov predicts that, if the geometry is right, his team should see the star dim on 7 May as the planet passes in front of it.

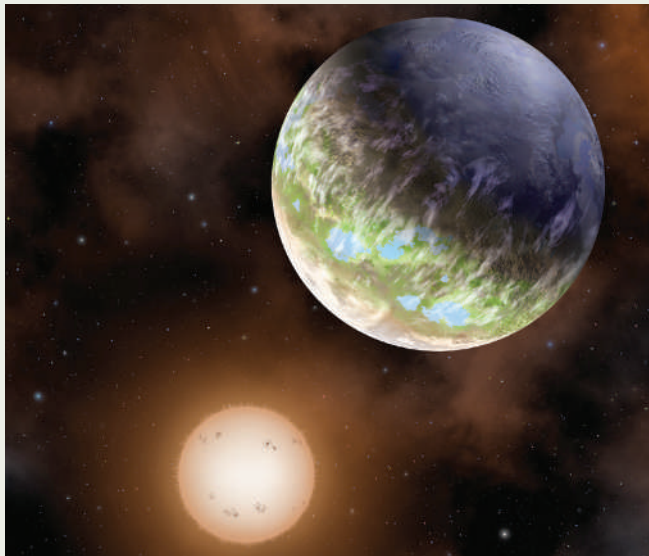
"If Gliese 581 c transits then the doors are open" says David Charbonneau, another Harvard exoplanet hunter. A transit would not only supply data about density, it would open up the possibility of follow-on observations that might reveal clues to the contents of the planet's atmosphere. A definitive failure to see a transit, on the other hand, would close off almost all lines of future observation with current

technologies. An exception is listening out for evidence of artificial radio transmissions. The SETI Institute of Mountain View, California, kept an ear out for signals from Gliese 581 in 1995 and 1997. "We didn't find any signal," says Seth Shostak, a SETI Institute astronomer. But the institute's next project — the Allen telescope — will nevertheless make looking at the Gliese 581 system a priority.

Without transits, the wild card of SETI is the only way forward on this particular planet. But Charbonneau is still uplifted by the discovery. He draws the inference from Gliese 581 c that there are more such low-mass planets around dwarf stars, and indeed Udry says he already has candidates for which he needs just a bit more data. It's not unreasonable to hope that within a few years enough data will have been found to make it likely that at least one will be observable in transit.

"One day we'll be lucky," Sasselov agrees. "And hopefully," he adds, "it'll be next week."

Katharine Sanderson ■



Planets in the habitable zones of red dwarfs could be strange new Earths.

L.R. COOK

Artefacts in ocean data hide rising temperatures

According to the myth — and a classic movie — Jason and his shipmates on the *Argo* suffered more than their fair share of reverses before bringing home the Golden Fleece. The international team of oceanographers working on the ambitious ARGO programme can surely sympathize. Instrumentation flaws have been discovered that undercut the programme's most striking finding to date. The problems have reinforced calls for caution when interpreting 'real-time' environmental data.

ARGO is an array of floats drifting freely around the oceans, 2,800-strong so far and still growing. Each float shuttles back and forth between the surface and the ocean depths measuring temperature and salinity, then sending data home via satellite when it is at the surface.

In 2006, data from the array led a team of scientists to the surprising conclusion that the world's oceans had cooled during 2003–05 — exceptionally warm years in terms of global surface temperature. The team published its findings in *Geophysical Research Letters*¹. Such apparent cooling was seized on by people keen to highlight the uncertainties in forecasts of global warming².

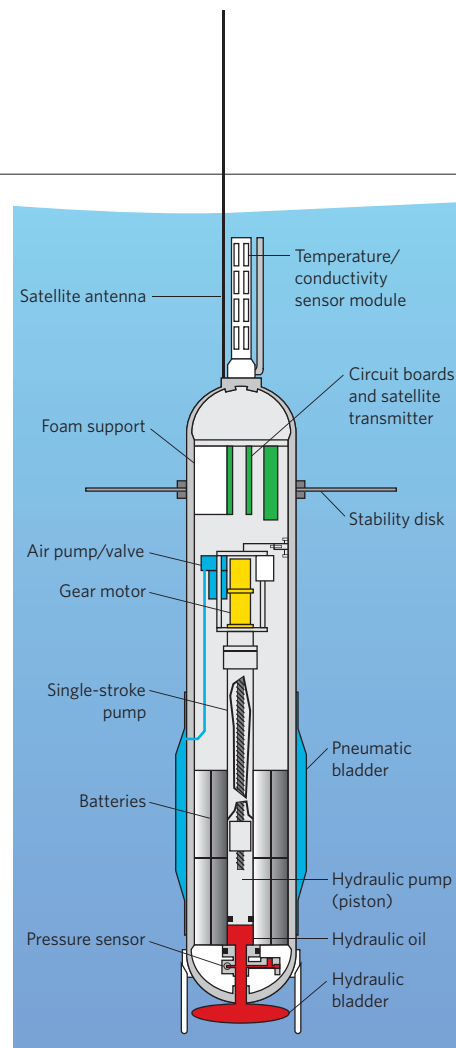
That cooling has now been shown to be an artefact. In some of the buoys — they are manufactured in separate batches — a software glitch caused the temperature and salinity data to be associated with the wrong depths. When the problem data are excluded from the analysis, the cooling trend drops below the level of statistical significance.

Novel observation technologies can often be error-prone. In 1997, for example, the discovery of a software problem involving NASA's TOPEX/Poseidon satellite forced a number of groups to lower previously published estimates of global sea-level rise³. And until three years ago, calibration problems with data retrieved from different satellites led to an apparently striking discrepancy between temperature trends at Earth's surface and in the atmosphere⁴.

The ARGO problem is a reminder that rigid quality control is vital in such cases, says Keith Alverson, director of ARGO's parent agency, the Global Ocean Observing System at the United Nations Educational, Scientific and Cultural Organization in Paris. The ideal is to be able to compare a new analysis with an independent data stream, he adds.

The flaw occurred in a batch of floats fabricated at the Woods Hole Oceanographic Institution in Massachusetts. It was detected when the temperature profiles generated by ARGO were compared with historical data from the regions where the cooling seemed to be most pronounced.

The authors of the original report, led by John Lyman of the National Oceanic and Atmospheric Administration in Seattle, Washington, submitted a correction to *Geophysical Research Letters* on 9 April; this had not been published when *Nature* went to press. Apart from the spurious ARGO data, they report a newly discovered bias in temperature profiles from expendable bathythermographs (XBTs)



The ARGO array of almost 3,000 floats measures ocean temperature and salinity worldwide.

that were also used for the analysis. Inexpensive XBTs have been a major data source for many oceanographic studies. The two problems concealed each other, says Josh Willis, an oceanographer at the Jet Propulsion Laboratory in Pasadena, California, and a co-author of the report: "When I first became aware of the problem I was really horrified."

Reform makes Italian research accountable

Italian research is headed for major reform — yet again. But this time it's for real, insists Luciano Modica, undersecretary of state for research, and the main architect of a reforming law that the Italian parliament is expected to pass by the end of May.

"The difference between this law and other attempts at reform is the element of autonomy," says Modica. For the first time, every research organization and

institute will have full control over its budget, recruitment and how it runs its research. At the same time, the institutes will be accountable to the government for their decisions.

Universities were given autonomy in 1990, but without much accountability. Now there will be budgetary consequences for any university or research institute that makes bad decisions. The new law

also creates an independent evaluation agency that will assess the quality of the research produced. And it gives the government power to withhold funds from institutes that do badly and to reward good performers with extra money.

"Not only money — also positions," says Modica, who this week created 1,600 positions for entrance-level professorships over the next three years.

These will be distributed to the universities ranked highest in Italy's first research assessment exercise, carried out in 2001–03. "The first ten universities in each field will get a certain number of these professorships, paid for by the government, and the others will get none," Modica says.

The decree also dictates that presidents of research agencies can no longer be simply named by the government, but will



CHIMP DENIED A LEGAL GUARDIAN

Court turns down request in case aiming for 'ape rights'.
www.nature.com/news

The ARGO Steering Team says that additional information sources should allow one-third of the biased profiles to be corrected exactly, and the remainder approximately. Until then, it is "not prudent" to use data from the ARGO array in combination with XBTs, the authors recommend in their correction. Other studies done using the data will need to be re-examined in this light.

Yet oceanographers remain excited about ARGO's chances of delivering true gold, saying it is the best thing to happen to the field in decades. "Factor it in and move on," says Gavin Schmidt, a climate modeller at the NASA Goddard Institute for Space Studies in New York. But he adds that scientists do need to be more aware of things that can go wrong with new observation systems — particularly if anomalous results turn up.

So has the ocean warmed after all? Yes, but not by as much as had been thought before the flaw in the XBTs was found, says Viktor Gouretski, an oceanographer at the Alfred Wegener Institute of Polar and Marine Research in Bremerhaven, Germany. Gouretski has reanalysed historical data, correcting for the bias in the XBTs⁵. According to this work, the top 3,000 metres of the oceans warmed by some 0.03 °C between 1957 and 1996 — around one-third less than when the bias was not taken into account.

The effect of the more recent bias in the ARGO data has yet to be worked out, however. "It's always a work in progress," says Schmidt. ■

Quirin Schiermeier

1. Lyman, J. M., Willis, J. K. & Johnson, G. C. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL027033 (2006).
2. *Nature* **442**, 854–855 (2006).
3. Nerem, R. S. *Science* **275**, 1049–1053 (1997).
4. Fu, Q. *et al.* *Nature* **429**, 55–58 (2004).
5. Gouretski, V. & Koltermann, K. P. *Geophys. Res. Lett.* **34**, doi:10.1029/2006GL027834 (2007).

be chosen from short lists drawn up by expert committees.

Among other things, this change is likely to lead to the ousting of Fabio Pistella, president of Italy's main research agency, the National Research Council. Constitutionally, such changes at the agency must be overseen by a temporary commissioner, not the incumbent president. A new president will then have to be appointed under the changed procedures once the commissioner's job is done. Pistella, who is not highly regarded, is unlikely to be on the post-reform short list. ■

Alison Abbott

India plans third Antarctic base

GOA

Sometimes it's nice to catch up with your old neighbourhood, even if you have been gone for some time. Indian researchers are taking this desire to extremes with a campaign for a new Antarctic base in Princess Elizabeth Land — a part of the continent that, 120 million years ago, was contiguous with India's eastern coast. But the idea has been opposed by other countries and environmental groups.

Last year, India's proposal for such a base was turned down at an Antarctic Treaty Consultative Meeting (ATCM) in Edinburgh. Now the proposal has been resubmitted to this year's ATCM, which started in New Delhi on 30 April. "This time we have made a strong scientific case for locating the base in Larsemann Hills," says Rasik Ravindra, director of the National Centre for Antarctic and Ocean Research (NCAOR) in Goa, India.

But the Antarctic and Southern Ocean Coalition (ASOC), a pressure group based in Washington DC that opposes the Indian base, continues to voice doubts. "There have been concerns about the integration of India's plans with the plan for an ASMA (Antarctica Specially Managed Area) by Australia, China, Romania and Russia," says James Barnes, ASOC's Executive Director. Ravindra says that India favours the Larsemann Hills being designated as an ASMA, and would accept the special controls on environmental damage that such a classification would entail.

A 75-page environmental evaluation report prepared by NCAOR seeks to allay fears that the Indian station would harm the Antarctic environment. The effect, if any, would be only "minor or transitory", the report concludes. India would

be willing to quit the base in 25 years, removing all signs of its existence, says Ravindra.

Some Indian researchers think that other countries are trying to exclude them. "It is like entering an unreserved compartment in a train to find the seats are taken," says Ravindra. They think they were in first and so others must keep out." Prem Chand Pandey, former director of NCAOR, suspects that the powers *in situ* have an economic motive: "The area is very rich in hydrocarbons and those who

Having permission refused for a permanent base will not stop India's research. "Nobody can stop us from going and doing research in the site we have chosen," says Maruthadu Sudhakar, a senior scientist at NCAOR who led a survey team to Larsemann Hills two years ago. But a new base would make the logistics easier. It would be closer to current work sites and it would be well sited for studies of the geomagnetic field — as the magnetic field lines studied at the Larsemann Hills would be those already



Home from home: India wants a research base in the Larsemann Hills.

are already in that area do not want others to come there." Although Antarctic resources cannot be exploited under the terms of the Antarctic treaty, that could change in the future.

One scientist notes that during Indian survey work in the area, helicopters from nearby Australian, Russian and Chinese bases were watching from above. These countries supported their opposition to the Indian base with the claim that the area had already suffered "human impacts." In a conciliatory move, Australia offered Indian researchers facilities at its station as an alternative to a new base. But India declined the offer because it "would not have been practical" and because Indians preferred to have their own base, Rajan says.

under investigation at magnetic observatories in Hungary and Denmark. The new base could also be a good ground station for the Indian ocean-observing satellite due to be launched in 2008.

But the main reason for selecting the site was to investigate how India broke away from Antarctica. Indian scientists hope that the Larsemann Hills will reveal vestiges of the Indian Mahanadi River, which flowed there about 130 million years ago when the continent was part of Gondwanaland. "The correlation between the two distant locations with respect to lithology, structure, tectonics and other geological constraints would help in fine-tuning the Gondwana fit," says Rajan. ■

K. S. Jayaraman

S. RAJAN/NCAOR

Missing gas saps plant theory

A team of plant scientists has cast doubt on one of the most startling research results the field has seen in recent years — the finding that green plants emit methane. Tom Dueck of Plant Research International in Wageningen, the Netherlands, and his colleagues say that they can find no evidence that plants produce the potent greenhouse gas.

Depending on how the comparison is made, methane's greenhouse effect is between 25 and 70 times as great as that of carbon dioxide. Until last year, methane was thought to be produced almost exclusively by microbes either in natural settings, such as swamps, or artificial ones, such as landfills, and in the guts of domesticated cattle. The claim by Frank Keppler of the Max Planck Institute for Nuclear Physics in Heidelberg, Germany and his colleagues that plants also emit methane, published in *Nature* last year (439, 187–191; 2006), threatened to overturn this microbes-only view of the global methane cycle. According to Keppler's team, plants might account for up to 40% of the 580 million tonnes of the gas that are emitted every year.

The claim was immediately controversial, not least because Keppler's team could not come up with a metabolic mechanism for their lab-based observations. One reason that the matter has not already been settled is that the amount of methane produced by each plant is low relative to the ambient methane concentration. To tackle this problem, Dueck and his

team grew plants in an atmosphere of 'heavy' carbon dioxide enriched in the isotope carbon-13 to give any methane produced a recognizable isotopic signature. But their studies detected no significant methane emissions (*New Phytol.* doi:10.1111/j.1469-8137.2007.02103.x; 2007).

Dueck acknowledges the difficulty of attempting to prove a negative. "The discussion is definitely not ended here," he says. "We are very curious to see how this is going to go. But if you are going to look for a mechanism [for methane emission], first you have to be able to measure it."

Both groups have criticized the other's choice of experimental method. Dueck says that Keppler's group kept plants in sealed plastic containers instead of flow chambers, and exposed them to sources of stress such as bright sunlight and high temperature, which could have produced methane as an artefact. Keppler retorts that the use of ^{13}C is an artificial piece of chemical trickery with unknown effects on plant metabolism, and also argues that methane production can vary by up to three orders of magnitude between species.

Nevertheless, without a mechanism, Keppler's claim remains open to attack. "We are aware that while some scientific groups are having difficulties in repeating our work, several others have been able to do so and their results agree with our findings," Keppler says. These confirmatory results have yet to be published.

Meanwhile, another plant scientist, David Beerling of the University of Sheffield, UK, says that he has not seen any methane using a method that is similar to Dueck's, but without relying on ^{13}C . His research has not yet been published, but Beerling thinks that if it joins Dueck's in the peer-reviewed literature, "the two together could kill off the theory".

The problem of how to explain puzzlingly high methane levels found above certain forests remains. But in light of the new results, atmospheric scientist Ed Dlugokencky of the Earth System Research Laboratory in Boulder, Colorado, thinks that the most likely explanation is "another, perhaps anaerobic source that we haven't identified yet".

But even if the theory can be verified, it is likely to have little bearing on regulations of man-made greenhouse-gas emissions, which are known to be just over half of the global total, Dlugokencky says. "Total emissions are known pretty well from the top down," he says. "If methane emissions are coming from forests then that probably just means less from wetlands."

Michael Hopkin



Plants might generate methane, but how?

3 GOOD REASONS

For NASA to feel smug

1 It has just helped Stephen Hawking (pictured) get his first taste of weightlessness, flying aboard a 'vomit comet' in preparation for a planned space journey next year.



NASA

2 The STEREO mission has captured the first three-dimensional images of the Sun, which should make for better forecasts of the weather in space.



3 Officials are expecting a visit from Britain's Queen Elizabeth II, who will drop in at the Goddard Space Center next week as part of a US tour. A flight on the vomit comet is not thought to be on Her Majesty's official itinerary.

R. LAUTENS/TORONTO STAR/ZUMA PRESS/NEWS.COM

SCORECARD

Victor Yushchenko
The Ukrainian president has pledged to rehabilitate those affected by the Chernobyl accident and to develop the territory contaminated by the blast 21 years ago...although he might have to wait for the radiation to dissipate first.

Anti-GM protestors
Red faces for British activists, whose meticulous plan to plant organic potatoes in a transgenic trial plot fell at the final hurdle — when they sabotaged the wrong field.

ON THE RECORD

"It's like a Thomas Hardy tragedy."

Massachusetts Institute of Technology's Leslie Perelman reacts to the resignation of the institute's dean of admissions Marilee Jones, who admits that she lied about her qualifications during her 28-year tenure.

Sources: NASA, Environment News Service, The Guardian, The New York Times

Denmark launches big push for protein power

A major centre dedicated to protein research is to be built in Denmark thanks to the largest ever donation to Danish research. The Novo Nordisk Foundation, which owns various healthcare and biotechnology companies, this week announced that it would provide 600 million kroner (US\$110 million) to fund the project at the University of Copenhagen.

A core component of the new centre will be a high-throughput facility to express and purify proteins, determine their structure and investigate their properties. The centre will focus on human disease, and will seek to formulate proteins for preclinical tests if they look promising as therapeutics. The university will keep the project's intellectual property, says vice-dean Birgitte Nauntofte.

Just as the study of genes scaled up into genomics in the 1990s, so in recent years protein researchers have been upgrading to proteomics. It is a substantial challenge: genes can code for more than one protein, and the products described in genes can be modified after being translated into protein. This means that although the human genome contains some 25,000 protein-encoding genes, a given person's various cells might use up to a million different proteins to do different things at different times in the course of a life. To crank up the complex-



Targeting proteins:
Søren Brunak (above)
and Matthias Mann.

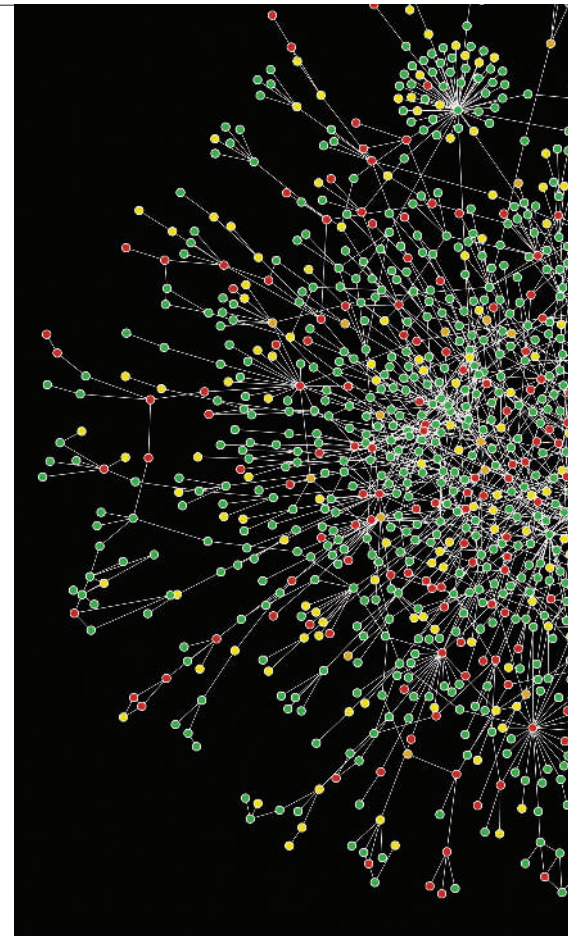


ity further, the proteins work in coordinated teams, requiring their relevant members to be in the right place at the right time if they are to generate the right response.

The Novo Nordisk Foundation Center for Protein Research will have five principal investigators, two of whom have already been appointed. One is Matthias Mann, a director at the Max Planck Institute of Biochemistry in Martinsried near Munich. Mann is a pioneer of protein mass spectrometry, a key proteomics technology. This technique, which smashes proteins into tiny fragments before reconstructing them by computer for identification and further study, benefits from the sort of large-scale computing power and bioinformatics capability

a large centre can provide. Mann will not quit his current position but will oversee independent groups working on biological mechanisms such as signal transduction and on stem-cell biology, particularly in relation to diseases such as cancer.

The other lead investigator so far named is Søren Brunak, from the Technical University of Denmark in Copenhagen. Brunak runs one of the largest bioinformatics departments in Europe. At the new centre he will focus on integrating what is known about creatures' physical characteristics with proteomic and genomic information in order to understand



how different protein complexes may relate to disease. "We'll have access to Scandinavian biobanks and electronic patient records renowned for their high quality," says Brunak, "and we'll find more by text mining scientific publication databases."

Mann and Brunak already collaborate in the Interaction Proteome Project, which is funded by the European Commission. "We've merged our skills before — for example we've created

Rapid sequencer puts virus in the frame for deaths

The discovery of a virus that may have killed three transplant recipients in Australia could mark a dramatic acceleration in the speed at which new pathogens can be identified. But it raises concerns that the ease with which such suspects can now be found could lead to researchers overlooking the need to firmly establish them as the cause of the disease in question. Fingering the wrong microbe could lead to inappropriate treatment or divert attention away from the real cause.

The three patients received organs from a single donor in Melbourne in December 2006. By January all three were dead. Ian Lipkin of Columbia University's Mailman School of Public Health in New York, who specializes in identifying new pathogens, heard about the case and collaborated with the Australian researchers who first took it on. After looking

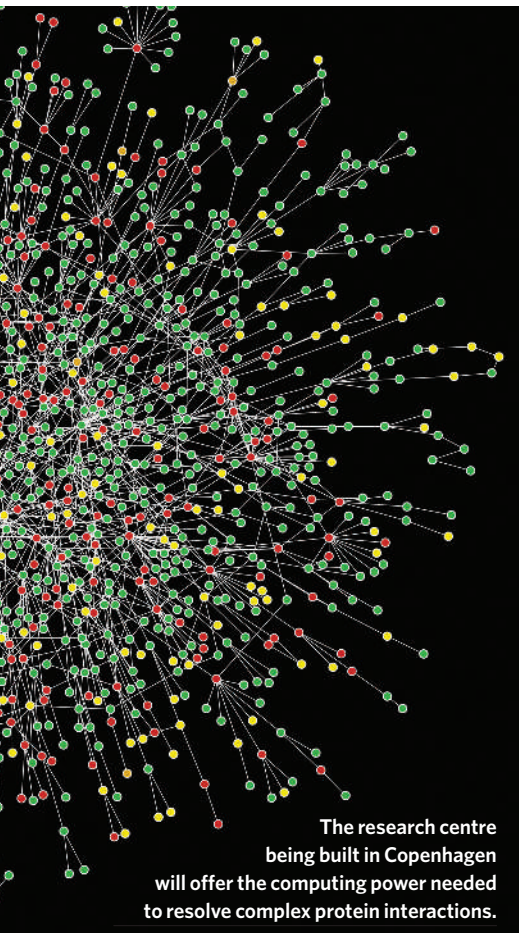


Transplanted organs can carry viruses.

for thousands of telltale signs of known pathogens without success, Lipkin, who sits on the advisory board of 454 Life Sciences in Branford, Connecticut, decided to use the company's technology to sequence genes from samples that had been filtered to enrich their non-human DNA component.

Roughly one month and

AI PHOTO/SPL



H. JEONG, UNIV. NOTRE DAME/SPL

what we call a 'temporal interactome' in one cellular compartment, the human nucleolus," comments Brunack, referring to studies of patterns in protein interaction studied over time as well as in space.

Building work for the centre will start immediately and it will open in 2008 with a hundred or so staff; some research projects will begin in other premises before this. ■

Alison Abbott

144,000 fragments of sequence later, a homemade algorithm in Lipkin's lab had pulled out 14 gene fragments that looked viral. Preliminary, unpublished analyses suggest they come from a new member of the family Arenaviridae. Although this sort of sequencing has been used to identify viruses in the past, the 454 technology cuts down on time and effort, says Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases, which funded part of Lipkin's work. Lipkin estimates that the technique could be used to process samples in a matter of days.

But simply finding a virus is not enough,

Live Earth taps into sense of joy

In some contexts, black suits over black shirts might seem depressingly funereal. But for Rob Reiner, Hollywood writer-director-producer-actor, and Al Gore, the Oscar-winning former US vice-president, the black-on-black look is the sort of chic more or less de rigueur when opening the Tribeca Film Festival. They were there because this year it features a series of shorts about climate change. And above the collars the two men were all about the joy, doing their best to give green activism a new, upbeat sell.

Their enthusiasm revolved around Live Earth, a global extravaganza set for 7 July that, Gore says, "is an opportunity for the whole world to come together at the same time in a spirit of joyfulness".

Live Earth is to be a media event in the tradition of 2006's Live 8, which focused attention on debt relief at the time of the G8 summit in Gleneagles, Scotland. It will include seven concerts on seven continents, and fits into Gore's stated plan to move the global culture past a "tipping point, beyond which political and business leaders and all sectors of civil society compete to offer policies and programmes that will sharply reduce emissions".

Naturally, nothing says

rock and roll like policies and programmes, and Live Earth has attracted a wide range of talent, although skewed, perhaps, to the tastes of Gore and Reiner's generation. At Tribeca, the band in the spotlight was Spinal Tap, whose life on the road was documented in Reiner's 1984 spoof rockumentary of the same name. The Tap, veterans of *Save the Ferret* and donors to the High Instep Foundation, are not only reforming for Live Earth but also releasing a new single, 'Warmer than Hell'.

A Reiner film of the reunion will be one of 60 short films on climate-change themes released at the time. Six of these shorts were shown at the festival, and ranged

from the dressed-up factual — an explanation of America's energy sources in the style of a rock video — to the strikingly metaphoric — a pretty face being spat on by passers by.

"All of these things that we are seeing where people are raising awareness are good," says Katie Mandes, communication director for the Pew Center on Global Climate Change, a policy shop in the Washington DC area. "You can't do the advocacy if you don't have heightened awareness, and we still have a long way to go." And she agrees with Reiner and Gore's cheerful stance. "I think it is important to start talking about solutions, to avoid the notion of hopelessness." ■

Emma Marris



None more green: Spinal Tap talk climate with Rob Reiner (left).

cautions microbiologist David Wang of Washington University in St Louis. Establishing that the virus actually caused the deaths is also critical, and is a lot harder. At present, the virus has been found in tissue from all three patients, and not in tissue from 60 controls. The virus's closest relative seems to be the lymphocytic choriomeningitis virus, which is believed to have killed organ-transplant patients in the past. But the team does not know whether the virus was also present in the donor. Lipkin says donor tissue samples have not been supplied to him by his Australian collaborators.

The guilt-by-association approach to pathogens can be misleading, warns microbiologist David Relman of Stanford University. Viruses do not necessarily behave in the same way as their closest known relatives. And the very fact that these new techniques work with raw sequence, rather than entities that could be grown and studied in the lab, makes follow-up experiments more challenging.

"It's not hard to find somebody you can implicate," says Relman, in police-procedural mode. "What's really hard is to nail the conviction." ■

Heidi Ledford

Fix for particle accelerator after magnets fail test

Physicists have agreed on how to fix a flawed set of magnets in the Large Hadron Collider (LHC), the world's newest particle accelerator. But it's still not clear whether the repairs will delay the full opening of the LHC, currently scheduled for mid-2008.

One of nine sets of 'triplet magnets', which will focus the machine's particle beams, failed on 27 March during testing. A review found that the magnets' support structures had been inadequately designed.

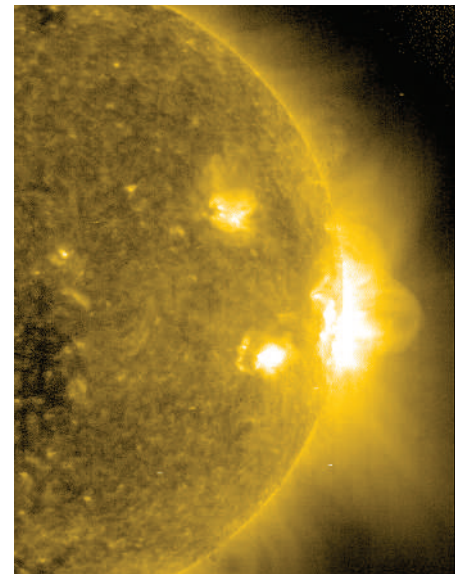
Researchers meeting at CERN, the European particle-physics lab near Geneva, Switzerland, on 24–25 April decided to place four metal cartridges around each magnet to reinforce its structure. These cartridges can be installed without bringing the magnets out of the LHC's tunnels, saving time and money. Further tests will take place in June.

Solar panel split over sunspot prediction

Experts charged with predicting the extent of the next maximum of solar activity have failed to reach agreement: issuing two separate forecasts for the number of sunspots expected in 2011–12.

Sunspots are linked to solar storms, which can disrupt satellite communications. A panel led by the Space Environment Center in Boulder, Colorado, split nearly evenly in its predictions for how severe the next cycle will be.

Six experts endorsed the idea that the cycle would peak with 140 sunspots in October 2011. Five others predict a maximum of 90 sunspots, in August 2012. The difference lies in whether you believe



SOHO/ESA/NASA

The next cycle of Sun storms is expected to be mild.

Living relative found for lonesome tortoise

Lonesome George, a Galapagos tortoise thought to be the last of his kind, may yet be saved from his loneliness. Genetic analysis of 27 tortoises on Isabela Island has uncovered a living relative — a male tortoise whose mother was from Isabela and whose father was from George's birthplace, the island of Pinta (M. A. Russello *et al. Curr. Biol.* **17**, R317–R318; 2007). This raises the possibility that other individuals with 'Pinta genes' live among the thousands of tortoises on Isabela. "Our priority now is to go back and look for more," says team member Jeffrey Powell of Yale University in New Haven, Connecticut.



J. R. POWELL

that the Sun retains a 'memory' from one solar cycle to the next.

But both predictions fall within the average range. "We're saying at the very least that we're not expecting a hugely active solar cycle," says panel chairman Douglas Biesecker, who is based at the Boulder centre.

US House backs bill to bar genetic discrimination

By a vote of 420 to 3, the US House of Representatives passed a bill on 25 April that would bar health insurers and employers from basing insurance coverage and job decisions on the results of genetic tests. Many US researchers support the bill, which they say will make it easier to recruit people for clinical trials.

The US Genetic Information Nondiscrimination Act is expected to come before the Senate later this month — where it has been passed unanimously twice before only for it to be blocked by the House. If the bill passes into law, the United States will join countries such as France and Austria in banning the use of genetic information in such decisions. Not all countries make this provision.

European grants give starters a helping hand

More than 9,000 young scientists have applied for 'starting grants' from the European Research Council, in response to its first-ever call for proposals. There are 200–250 grants available, worth around €290 million (US\$400 million) in total.

The grants are to help young researchers in Europe make the transition to becoming

group leaders. Low success rates at securing grants have in the past deterred scientists from applying for European Union funds, so the council's awards will be decided through a two-step evaluation process.

In the first round, applicants submit only an outline of their research proposal. Shortlisted candidates, selected by July, will be invited to submit more detailed proposals.

Oversubscription of research programmes is also a problem in the United States, where officials at the National Science Foundation, for example, would like to boost that agency's application success rate to 25% from 21% in 2006.

Arnold issues ultimatum on carbon emissions

Watch out — the Governor has issued an ultimatum. California governor Arnold Schwarzenegger has given the US Environmental Protection Agency's administrator Stephen Johnson six months to approve the state's request to be allowed to regulate the carbon dioxide emitted by cars and trucks, or he'll sue.

California must get permission to regulate from the agency under the Clean Air Act. Johnson had been stalling while the agency pursued in the courts the position that greenhouse gases did not count as 'pollutants' and could therefore not be regulated under the act. The Supreme Court shot down this position on 2 April.

At a Senate hearing on 24 April, Johnson faced intense questioning on how and when the agency would start regulating carbon dioxide. He repeated the phrase "expeditiously and responsibly" nearly a dozen times.

BUSINESS

Patenting the obvious?

A patent ruling has delivered a serious blow to one US state's hopes of a biotechnology boom. **Erika Check** reports.

A preliminary decision by the US Patent and Trademark Office to revoke three stem-cell patents is being viewed as a victory for stem-cell researchers — except in Wisconsin, the state that held the patents.

The researcher and advocacy groups that challenged the patents say they hope their success will spark a wider debate on whether too many sub-quality biotechnology patents are being issued in the United States.

"A serious, public policy discussion needs to be had about what researchers can patent," says John Simpson of the Foundation for Taxpayer and Consumer Rights in Santa Monica, California, and one of two non-profit organizations that challenged the patents.

The patents in question covered methods for making embryonic stem cells from primates, including humans, developed by James Thomson of the University of Wisconsin-Madison. They are held by the Wisconsin Alumni Research Foundation (WARF), a charity that helps scientists to exploit their inventions commercially.

The patents have been a linchpin of efforts, led by state governor Jim Doyle, to promote Madison as a biotechnology hub. But government efforts to attract biotechnology investment are notoriously unreliable, and the patents offered no guarantee that Wisconsin would succeed.

After the patent office published its decision last month, Doyle sought to play down its probable effect. "Wisconsin is and will remain at the forefront of stem-cell research for many years to come," he said.

The WARF patents have long irked stem-cell researchers, who asserted that the terms of the licences have stifled academic and entrepreneurial innovation (see *Nature* 435, 272–273; 2005). But it was a simmering conflict with the California Institute for Regenerative Medicine (CIRM) that got the patents into trouble. Last March, a WARF official told a conference of biotechnology entrepreneurs that WARF expected the CIRM to pay it licence fees and royalties. This January, the foundation backtracked from that statement, saying that the CIRM wouldn't have to pay the fees. But the

issue had already stirred up a hornet's nest in California.

Researchers such as Jeanne Loring, a biologist at the Burnham Institute for Medical Research in La Jolla, took exception to WARF's stance. Last July, Loring joined forces with two pressure groups, the Foundation for Taxpayer and Consumer Rights and the Public Patent Foundation in New York, to ask the patent office to re-examine the patents. The challengers argued that human embryonic stem cells had been isolated before, and that Thomson had not acknowledged this in his applications.

In its 2 April statement, the patent office said that it accepted these arguments, and intended to revoke the patents. WARF has until June to respond to the decision, and if it is unhappy with the outcome, it can then initiate an appeal. The patents will be treated as valid until the re-examination process is complete — that is, until WARF's response and the possible appeal have concluded. That could take years.

In the meantime, WARF holds three unchallenged patents on embryonic stem cells and has applied for dozens more. And Tom Still, president of the Wisconsin Technology Council, points out that stem-cell money constitutes less than 1% of WARF's total licensing revenues. "Life will go on," says Still.

But Wisconsin's image has undoubtedly taken a hit. Some in the state see the episode as an unfair assault on its achievements, launched on shaky ground. "These patents are not stifling innovation," Still argues. "WARF has done a lot to make stem-cell lines available to researchers at a very reasonable price."

The challenge is putting renewed pressure on Governor Doyle's drive to establish a biotech industry in Wisconsin.

At last year's Biotechnology Industry Organization conference in Chicago, for instance, the state spent \$270,000 on a booth to sell itself to potential investors. The booth touted Thomson's work, as well as a programme of tax credits for innovative businesses and a \$375-million Institutes for Discovery at Madison as the centerpiece of a \$750-million plan for stem-cell



Stem-cell research in James Thomson's lab has generated dozens of patent applications.

research that he announced in 2004.

Critics of the patents have sought to assure Wisconsinites that they have no ill intentions towards the state. "I'm sorry to see people taking it that way, and that's not what's at stake here," says Simpson.

Early success

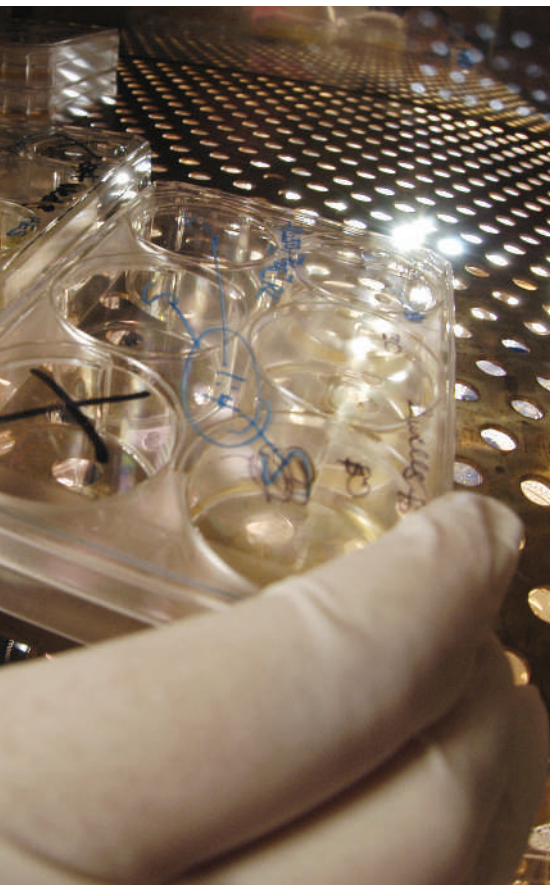
And, the critics argue, the patent office's announcement is already benefiting researchers. "WARF was asserting its case in an aggressive way, and it has had its wrist slapped. That's a benefit to all researchers in the United States," Simpson says.

Jeanne Loring adds that researchers are now thinking about entering a field that they may have shied away from before because of the aggressive management of the patents. "The scientists I have spoken to feel like some weight has been lifted from them," Loring says. "It's not so much the amount of a fee — it's the uneasy feeling that you are going to be violating some law at some point if you continue in this field."

Simpson and his co-challengers have been emboldened by their early success. They sug-



"Wisconsin is and will remain at the forefront of stem-cell research for many years to come." — Jim Doyle



J. MILLER

gest that what's really at stake is a larger question about the quality of patents granted in the United States. In an effort to glean money from licensing fees, they argue that university technology transfer offices are aggressively patenting researchers' work, even though it may not be as novel as they claim.

Those challenging the patents hope that this case will have an effect on technology transfer offices across the country. "Those technology transfer officers that are sitting on a bunch of bogus patents are shaking in their boots, and they should be," claims the Public Patent Foundation's Dan Ravicher.

Simpson adds that there is a broader movement underway to tighten the patent office's standards. A crucial Supreme Court case on the question of 'obviousness' — the yardstick used to judge patentability — was argued before the court last November. A decision in that closely watched case, between KSR International and Teleflex, has not yet been handed down.

Simpson says that the case is a sign that the larger public discussion about patents is slowly gathering steam. And its effects could ripple across the country — especially in the biotechnology industry, which depends on patents for its very existence. ■

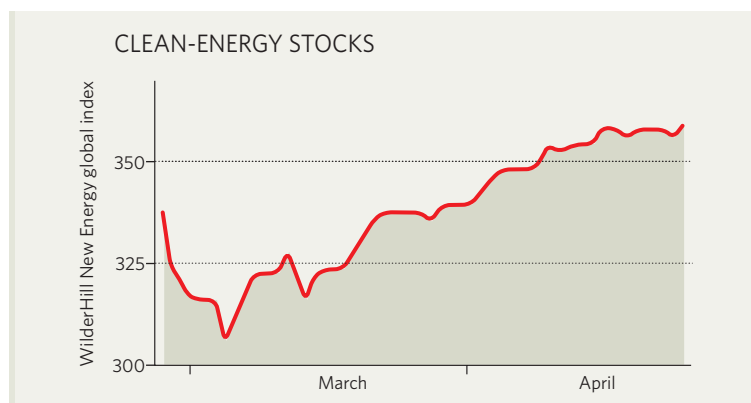
IN BRIEF

BIOTECH PLUNGE British drug firm AstraZeneca said it would pay a cool US\$15.6 billion to acquire MedImmune, a biotechnology company based in Gaithersburg, Maryland, whose main product is a nasal-spray flu vaccine. The deal demonstrates the extraordinary lengths that major drug companies are now prepared to go to in order to get their hands on promising biotech, observers say. Last year, MedImmune, which was set up in 1988 by Wayne Hockmeyer, a biologist who had worked at the Walter Reed Army Institute of Research in Maryland, declared profits of \$75 million on sales of \$1.3 billion.

PLEASING VISTA Microsoft reported soaring profits for the first quarter of this year, dispelling doubts about the commercial success of its Vista operating system. The computer-software firm reported earnings for the first three months of this year of \$4.9 billion — up 65% on the same period in 2006 — on sales of \$14.4 billion. US sales of computers operating on Microsoft software surged forward by two-thirds the week after Vista was launched, despite sniffy press reviews.

SHARES SUSPENDED Oxonica, the British nanotechnology company, temporarily suspended trading in its shares on 27 April, after receiving notification from a Turkish oil firm that a \$12-million contract to supply the oil firm with fuel additives had been cancelled. The contract was ended a month after Oxford-based Oxonica admitted that the results of tests for the oil company had been 'inconclusive' (see *Nature* 446, 963; 2007). Oxonica is locked in legal dispute with Neuftec, a company registered in the Caribbean island Dominica, over the licensing of a cerium-oxide fuel additive.

MARKET WATCH



Clean-energy stocks have soared since early March on the back of a general market upturn and continued market buzz about alternative energy sources. But analysts warn that the optimism is becoming excessive.

The WilderHill New Energy Global Innovation Index (symbol NEX on the American Stock Exchange), which measures stocks in companies with interests in energy sources other than fossil fuels or nuclear power, hit its highest-ever level as investors poured money into the fashionable sector.

Michael Liebreich, chief executive of London-based New Energy Finance, which compiles the index jointly with WilderShares of California, says that generally buoyant sentiment is behind the index's upward march.

"There has been almost no bad news for clean energy," he says. "Sentiment is basically running away with itself and it is due for a correction. At some point in the next 12 months, there will be one."

The point is reinforced by a report on 'clean technology' issued this week by New York-based Lux Research. "The warning signs of a bubble are appearing in the energy segment," the report says.

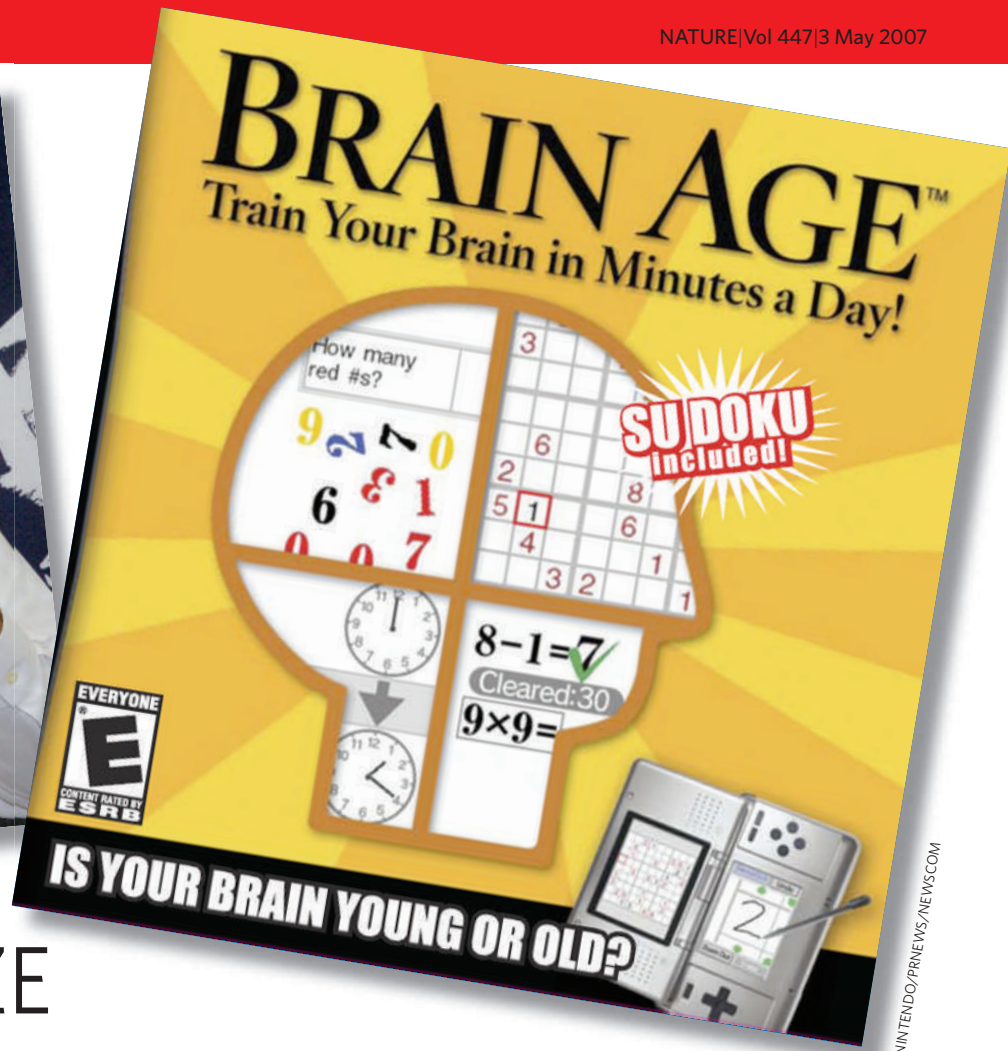
Liebreich notes that the positive sentiment is being driven by strategic events such as Californian legislation last September, which has set ambitious long-term targets for carbon emissions (down 80% from 1990 levels by 2050), and Norway's 19 April pledge to go 'carbon neutral'. "It's a small country but a big step," he says of Norway's plans.

The stand-out divisions of the clean-energy sector are wind and solar power. However Liebreich notes that some solar-energy companies, desperate to secure supplies of silicon ingots or wafers, have locked themselves into expensive, long-term contracts that will come back to bite them later on, as new supplies come on stream and prices of these commodities fall. ■

Colin Macilwain



KYODO NEWS



NINTENDO/PRNEWS/NEWS.COM

BRAIN CRAZE

Neuroscientist Ryuta Kawashima promotes the idea that computer games can boost the ageing brain — but others in the field remain sceptical. **Ichiko Fuyuno** investigates.

On a chilly, rainy day last November, Ikuyo Narawa climbed on a tour bus in Tokyo in the hope of revitalizing her brain. She knew that she drank too much and never exercised, and she thought that her memory was weakening. So, along with 40 companions, she travelled to the scenic Tateshina highland in Nagano Prefecture, where she tried hiking and handicrafts, ate organic meals and bathed in hot springs.

At the start of the tour, the 37-year-old dental assistant took several simple computer tests, such as clicking a mouse as soon as a lamp lit red, and was surprised to be told that her 'brain age' was 61 years. After two days of activities, she took the test again and found that her result was, disappointingly, unchanged. But on her fifth time, attempting the test at home, her brain age dropped to 20. "I was very relieved," she says.

The brain-train tour, organized by the region's local chamber of commerce and a travel agency, is the latest example of a brain-exercise boom that is consuming Japan and spreading across the globe. In Japan, TV quizzes aimed at boosting the brain are popular, and fashion magazines run rejuvenating advice for the mind

alongside that for the skin. And throughout the world, a proliferation of books, websites and software is available claiming to preserve our mental capabilities.

At the centre of this craze is a ragingly popular video game developed by Japanese video game maker Nintendo and based on the research of Ryuta Kawashima (pictured above), a neuroscientist who specializes in brain imaging at Tohoku University in Sendai. Players of the game — called 'Brain Age' in the United States and 'Dr Kawashima's Brain Training' in Europe — use a console to complete simple tasks such as reading aloud, multiplication and memorizing words. In promotional material for the game, Kawashima says that daily training on these activities can "help to prevent a decrease in brain function". He promotes the

idea that these types of activity enhance blood flow to the brain's prefrontal cortex — the region of the brain that regulates aspects of memory, reasoning and some of the other complex behaviours that deteriorate with age.

Many neuroscientists and gerontologists are sceptical of the claims made for Brain Age and similar games, saying that there is scant evidence that any type of brain exercise can halt

mental ageing. Kawashima has published a paper¹ suggesting that the type of mental activities in the game can help elderly people with dementia, but critics say that these findings do not necessarily apply to healthy adults. A player's score could well improve with practice, but whether that translates into an improvement in other mental tasks or everyday life skills remains unknown. "That's what I think is a big question," says Timothy Salthouse, a researcher of cognitive ageing at the University of Virginia, Charlottesville. "I don't think there is scientific evidence that the improvements after mental exercise can be generalized beyond what you have trained on."

In a rapidly ageing population, people are understandably keen to live a long life with their mental capacities intact. Neuroscientists think that from as young as 30, changes in the chemistry and connections between nerve cells cause some cognitive functions to decline. Brain-training programmes typically claim to slow this mental descent or even to recover some lost ground. Some studies have shown that physical exercise can prevent some of the brain's deterioration (see 'Body and mind'), and the idea that simple mental activities could do the same is enticing.

Brain training has already earned Kawashima celebrity status in Japan. He studied to be a

"People are paying for Brain Age. They deserve to know whether it really works"

— Dorothy Bishop

physician but, after learning about imaging techniques at Sweden's Karolinska Institute in Stockholm, he became one of the first to pursue brain-imaging research in Japan. Using techniques such as functional magnetic resonance imaging (fMRI), he showed that blood flow to the brain's prefrontal cortex increases when people are calculating sums quickly or reading aloud. He has published more than 100 drill books with such exercises for adults.

Mental work-outs

In 2001, Kawashima won a coveted ¥60-million (US\$500,000) grant from the Japan Science and Technology Agency in Saitama to study whether mental exercises improve cognitive function in elderly people. He and his team investigated 32 individuals who had been diagnosed with Alzheimer's dementia in Eiju-no-sato nursing home, Fukuoka. Over six months, half of the people were asked to do simple calculations and language tests, such as arithmetic division or reading fairy tales aloud; the other half received no training. Kawashima's team measured their cognitive status before and after the training with two widely used tests to diagnose dementia — the Mini-Mental State Examination (MMSE) and the Frontal Assessment Battery (FAB). The tests included questions such as "What day of the week is it?" and "What do bananas and oranges have in common?"

People in the training group improved their FAB score, maintained their MMSE score and became more communicative and independent



The Brain Age game claims that its cognitive exercises can help the ageing brain.

NEWS.COM

than they had been before the training. The control group, however, showed no change in FAB score and a decline in the MMSE score¹. In addition, a 75-year-old man in the treatment group regained his ability to go to the toilet by himself, and a 77-year-old woman who used to come to the learning centre in her pajamas started to arrive fully dressed.

Kawashima and the other authors acknowledge the study's limitations. They could not

tell whether the cognitive improvements were attributable to the training itself, or to the extra attention and social interaction the individuals received from the experimenters and nursing staff. But Kawashima thinks that the method may have stimulated the prefrontal cortex, causing improvements in general cognitive functions such as communication. He called the method 'learning therapy', and it has now been introduced at 300 nursing homes across Japan at a monthly cost of ¥1,575 per user. Kawashima has conducted two more unpublished studies of the same training on elderly people with and without dementia and says that the training improved their MMSE scores.

Adult entertainment

At the end of 2004, Kawashima was contacted by Nintendo. The company thought that his drill books could be turned into a stimulating game that would attract those adults who usually shy away from conventional video games.

Working with Nintendo to develop the game, Kawashima says that he studied 120 Japanese people aged from their 20s to their 70s. He used a technique called optical coherence tomography to examine blood flow in their brain while they tackled dozens of exercises, such as adding up numbers and memorizing Chinese characters. The tomography technique is not as accurate as fMRI, but Kawashima says that he used it because it was quicker and easier for the volunteers. He and his research team selected 15 tasks that boosted blood flow to the prefrontal cortex and calculated an average score for each age, which they used as a basis for the Brain Age game.

BENELUX/ZEFA/CORBIS

Body and mind

Physical activity is already heralded for warding off cardiovascular disease and various other conditions. A growing body of evidence suggests that it can improve mental health too.

In a recent study, 11 young people were given a 12-week aerobic exercise programme. Afterwards, the participants performed better at memory tests. The exercise also promoted blood flow to the dentate gyrus, a region of the brain important in memory and cognitive ageing, and the authors suggest that it promoted the birth of new neurons there (A. C. Pereira *et al. Proc. Natl Acad. Sci. USA* **104**, 5638–5643; 2007). But researchers are not sure

exactly how physical exercise benefits the brain, or how much exercise is best for brain function.

Everyday life requires numerous skills, points out Warner Schaie, a psychologist studying cognitive development at Pennsylvania State University in Philadelphia. So elderly people should continue to do intellectually stimulating activities and to interact with other people, he says. The activities can be tailored to individuals' taste — from volunteer work to square dancing and travelling.

Many experts say that leading a generally healthy lifestyle — including physical



exercise, a good diet and mental and social activity — is probably the surest route to cognitive health. "The hypothesis we have is that mental and physical activities, social engagement and lowering vascular risks can lessen your chances of a decline in cognition," says Marilyn Albert, from Johns Hopkins School of Medicine in Baltimore, Maryland. "But we have a long way to go to understand that." I.F.

The video game debuted in Japan in May 2005. Sales of the ¥2,800 product, together with its sequel, have reached more than 3 million units in Japan — in a market in which 1 million units is considered a hit. The game was introduced in the United States, Europe and Australia in 2006 and in South Korea earlier this year. Kawashima says that he uses all of the royalties from Nintendo and other companies — more than ¥400 million in 2006 — on his research, including construction of a new laboratory near his office.

Lost in translation

Other neuroscientists, many of whom say that they respect Kawashima's work, express discomfort with his Brain Age concept. Although the game could work in principle, they question whether simple mathematics and reading tests of the type in Brain Age are any more effective than other cognitive tasks at boosting blood flow to the prefrontal cortex. And they say that there is little evidence that a brief boost in blood flow would improve brain processes or the everyday skills that decline in normal ageing adults. "I see no reason to believe that Brain Age gains will transfer to other kinds of cognition, or to real-world function," says Michael Marsiske, who also studies interventions to improve cognitive performance, at the University of Florida, Gainesville.

Earlier this year, 41-year-old Marsiske played the game himself. He says the game was entertaining and fun and that his brain age dropped from 78 to 26 after three days. But much of the improvement could be attributed to practice, he says. "Users may get the illusion of huge gains when starting with Brain Age, but these have more to do with learning the device than actual mental improvements."

Marsiske and other neuroscientists say that they would like to see well-controlled, published studies to show that Brain Age benefits those who are buying it. "People are paying for it. They deserve to know whether it really works," says Dorothy Bishop, a neuroscientist at the University of Oxford, UK. Kawashima is indifferent to this criticism and says that he does not intend to conduct more detailed studies on the game's effects because his earlier study on those with dementia, and his additional unpublished work, demonstrated that his learning therapy works. "I am sure it works because all the data we obtained so far have been showing positive



Grey gaming: researchers are testing whether computer exercises can improve elderly people's skills.

results," he says. "The most important thing is that behaviours of older people have got better with our training method." Yasuhiro Minagawa, a spokesman for Nintendo, says that the company is not in a position to comment on the scientific evidence behind the game, and the company is confident that Brain Age provides high-quality entertainment.

Other scientists are eager to test more rigorously whether mental work-outs can enhance the ageing brain. "A lot of people are interested in this area and are working hard to see whether or not

this kind of short-term training has a long-term benefit," says Marilyn Albert, an expert on Alzheimer's disease at Johns Hopkins School of Medicine in Baltimore, Maryland.

A firm called Posit Science in San Francisco, California, is conducting research into Brain Fitness Program, its best-selling computer-based exercise (costing from \$395), in which users distinguish similar sounds, reconstruct sequences of words and do other mental exercises. A team led by Michael Merzenich, the company's chief scientific officer and a researcher on cortical plasticity at the University of California, San Francisco,

assigned 182 participants aged over 60 to one of three groups. The first group performed brain exercises on a computer, the second watched and listened to educational DVDs on their computers and the third had no computer time. After the 8–10 weeks of the study, the training group had improved at the tasks in the program and in other standardized assessments of memory it had not been trained in², whereas those in the control groups showed no improvements.



"I see no reason to believe that Brain Age gains will transfer to real-world function."
— Michael Marsiske

Perhaps the best study on mental training so far, Albert says, was that published by a team including Marsiske late last year³. It suggested that healthy elderly people given a cognitive work-out can gain long-lasting mental benefits, which may have transferred into improvements in their daily living activities. More than 2,800 mentally healthy adults aged over 65 were randomly assigned to receive ten hour-long training sessions over five weeks or to a control group that received no training. The team found that the training groups had improved cognitive ability for the specific tasks they had trained on, and that the benefit lasted for as long as five years. And compared with the untrained controls, participants in one of the training groups reported that they had less difficulty in performing routine tasks, such as preparing meals and using the telephone.

Many neuroscientists are optimistic that brain training will have proven benefits. And they see no more harm in doing computer exercises than in completing brain teasers such as crosswords or sudoku, and can at least give a warm glow of accomplishment. At the Eijuno-sato nursing home, those with dementia are given basic exercises in reading and calculation for 15–20 minutes a day. Head of the nursing home Ritsumi Yamasaki says that this schedule has improved patients' behaviour more than therapies she had tried before, such as gardening and karaoke, and it helps their interaction with carers.

"I tell people to do it if it's enjoyable," agrees Salthouse. "There's little evidence that it's damaging or harmful, and we may eventually find out there are some benefits."

Ichiko Fuyuno is a reporter for Nature in Tokyo.

1. Kawashima, R. *et al. J. Gerontol. A* **60**, 380–384 (2005).
2. Mahncke, H. W. *et al. Proc. Natl Acad. Sci. USA* **103**, 12523–12528 (2006).
3. Willis, S. L. *et al. J. Am. Med. Assoc.* **296**, 2805–2814 (2006).

DEEP DIVISIONS

Archaeologists are unearthing remarkable finds in Jerusalem. But the digs have sparked an argument over who should run the site and present the results to the public. **Haim Watzman** reports.

In the ancient heart of Jerusalem, the one-kilometre-square walled area known as the Old City contains some of the most sacred sites for three religions, and as such it is no stranger to religious and political conflict. Yet some of the latest disputes are centred just outside the massive walls, in a Palestinian village known as Silwan that is now a neighbourhood within greater Jerusalem.

Here, archaeologists are battling over the interpretation of major ongoing excavations. The site, known to Israelis as the City of David, lies under part of Silwan and is operated by a Jewish settler organization. Some Israeli archaeologists are openly critical of this organization's aims, while many Palestinians claim that the digs are damaging their property. At the heart of the debate is the question of who should be allowed to control the site, oversee excavations, and present the findings to the public.

The part of Silwan that lies on top of the site contains some four dozen homes of Palestinian Arabs, and 20 homes of Israeli settlers. The houses stand on the ruins of centuries of Muslim and Byzantine habitation, which in turn cover the Jerusalem that was sacked and burned by the Babylonians in 586 BC, and before that besieged by the Assyrians in 701 BC. Farther down lies evidence that might help confirm — or refute — the Bible's account of a prosperous united Israelite kingdom in the tenth century BC, ruled by the kings David and Solomon. And deeper still are the remains of an even more ancient city that prospered around 1800–1700 BC, during the Middle Bronze Age.

So it's hardly surprising that archaeologists have been excavating here for the past 140 years. Digs are now under way at several sites in Silwan — both between and beneath the homes of the village's inhabitants. Each month, busloads of Israeli schoolchildren and Israeli and foreign tourists flock to see the unearthed finds.

But to some, the way in which these finds are presented and explained to the visitors is a major cause for concern. The visitors' centre

at the site is run by a non-profit foundation called Ir David (which means 'City of David' in Hebrew). The group was established in 1986 to promote excavation and tourist development of the site, and is popularly known by its Hebrew acronym, Elad. The foundation often helps to organize funds and support for archaeological excavations at the site.

But Elad also has another goal: promoting Jewish settlement in the village of Silwan. To that end, it has reclaimed formerly Jewish houses, evicting the Palestinian residents and replacing them with Jews, and has purchased Palestinian houses — sometimes using means that its Palestinian and Israeli critics charge are legally questionable. Renovation of these homes and development work for residential and tourist purposes have necessitated salvage excavations that have inconvenienced the Palestinian residents and sometimes caused damage to their property.

"We can't run any site if the residents don't want us."

— Yigael Ben-Ari

Elad's members and supporters are nearly all nationalist Orthodox Jews who believe that Jewish settlement in the territories captured by Israel in the Six-Day War of 1967 is God's will and a precondition for the arrival of the Messiah.

Although archaeologists digging at the site say that Elad has not pressed political interpretations on their work, some of them have raised concerns that the organization's religious and political goals are incompatible with the role of running a national park containing an important archaeological site. Indeed, they charge that Elad is using its position to promote a distorted version of history — merging myth and legend with archaeological fact.

This apparent conflict of interest has prompted a group of archaeologists to initiate legal moves to get the Israeli government to take control of the site from Elad. "They are taking over public land," says Rafi Greenberg, an archaeologist at Tel Aviv University who is one of the organizers of the initiative.

But Ronny Reich, an archaeologist at the University of Haifa who has excavated at the City of David, notes that Elad does not try to dictate who can work there. "I don't think anyone can tell a group of people that wants to participate in the excavations that they can't," he says.

Excavations near this ancient tunnel have revealed fresh details about how the Gihon spring was used.

M. RICHARDSON/CITY OF DAVID



Just below Jerusalem's city walls stretches a hill that is home to the archaeological remains known as the City of David.

Elad's spokesman refused *Nature's* request to interview a representative of the organization and asked that questions be submitted in writing. When they were, the spokesman failed to provide answers, despite repeated promises to do so.

Fact or fiction?

One way to experience the Elad view of the City of David is to tour the site with an Elad-trained guide. It is possible to visit the excavations on your own or with a guide you've brought yourself. But the default option for tourists and school groups is to hear the narrative that asserts the Jewish claim and historical connection to the site, say Greenberg and his colleagues.

There is some truth to these claims, as a *Nature* visit to the site suggests. The tour guide provided by Elad was well-spoken and knowledgeable, but mixed myth and fact in her presentation. For example, she asserted that the reason David chose the site for his capital is that it lies just below the Temple Mount, which is identical to Mount Moriah, the site where, according to the Bible, Abraham took his son Isaac to offer him as a sacrifice to God. Although the identification of the Temple Mount with Mount Moriah is well-established in Jewish tradition, there is no archaeological evidence for Abraham's presence on the site — or indeed for the existence of Abraham and Isaac.

In fact, a handful of archaeologists go so far as to say that David and Solomon may also be largely mythical characters. This view is rejected by most experts on the period — they tend to agree that it is likely the two ancient rulers did reign in Jerusalem. But many scholars argue that the evidence discovered so far — both at the City of David and at other sites in the region — indicates that the biblical description of the extent

and wealth of their kingdoms is exaggerated.

Furthermore, the Elad guide made no mention of Byzantine and Muslim settlement, giving the impression that the site is solely a Jewish one. But this may not be too surprising, given that she had only about an hour to explain the site and that her audience consisted of Israeli Jews, including a number of easily bored children and teenagers. Under those conditions, the presentation of any archaeological site would no doubt be geared more towards storytelling than to the detailed technical facts of what the archaeologists have found and how they interpret the evidence.

Few dispute that this complex site has yielded some major discoveries in recent years. Near the top of the hill in Silwan is an ongoing excavation led by Eilat Mazar of the Hebrew University and sponsored by Elad, the Shalem Center (a Jerusalem-based research institute), the Israel Antiquities Authority and the Society for the Study of the Land of Israel and Its Antiquities. In early 2005, Mazar's team uncovered a large stone structure, and dated pottery found inside the structure to early in the Iron Age IIa

period (around 1000 BC), which corresponds to the time of King David. So far, several large rooms have been uncovered, as well as walls two to three metres wide. In March, Mazar announced the discovery of another 20-metre section of the structure's outer wall, further evidence of its huge size. She believes that the massive nature of the structure indicates that it must have been an important public building.

And because it is located close to the Temple Mount and at a commanding position in the city, she believes that it is the palace that, according to the Bible, David built after conquering Jerusalem and making it his capital in the early tenth century BC¹.

Age concerns

But some archaeologists dispute her dating and interpretation. Israel Finkelstein of Tel Aviv University, for example, is a leading proponent of the view that many archaeological remains throughout Israel dated to the early tenth century BC — the time of David and Solomon — are actually nearly a century younger. On the basis of his later dating of the artefacts in question, and of the lack of references to a large Israeli kingdom centred in Jerusalem in the records of near-Eastern cultures, he argues that the rulers of Jerusalem were not significant players on the international stage until much later. "Mazar has done fine and important work," he says, "but interpretation is another matter. The structure she found can't be dated unambiguously."

Finkelstein thinks that the large stone structure instead dates to the ninth century BC — the period in which Omri and his son Ahab organized the northern Israelite tribes into a powerful kingdom. He argues that the





Palatial digs: Eilat Mazar (inset) believes the large building she uncovered in Silwan in 2005 is the palace of King David.

influence of the more powerful kingdom to the north, to which the rulers of Jerusalem were probably vassals, may have been the spur for major construction. Or perhaps, he suggests, the collapse of Omri's dynasty in the mid-eighth century BC may have created a power vacuum that the kings of Jerusalem were able to fill.

Another collection of artefacts, found in 2005 near the Gihon spring at the bottom of the Silwan hill, also touches on this debate. Reich and Eli Shukron, who conducted salvage excavations for the Israel Antiquities Authority around the spring, discovered a large number of bullae — clay seals placed on ancient letters. The researchers dated the bullae to the late ninth or early eighth century BC. The figures on some of the seals are Phoenician, a sign that Jerusalem was trading with the coast at the time. Further evidence is provided by some 10,000 fish bones found with the bullae — around 90% of which come from saltwater fish.

Tunnel vision

Ironically, although the Gihon spring was long the city's sole source of water and is mentioned in the Bible, it attracted relatively little archaeological attention until Reich and Shukron began digging there. Reich thought it so unpromising that he initially resisted being assigned to dig there by the antiquities authority, his employer at the time.

"Eleven or twelve years ago Elad received a permit from the municipality to build a visitors' centre by the spring," says Reich. "The head of the antiquities authority, Amir Drori, told them that, by law, they had to fund a salvage excavation at the site before building."

Reich and Shukron's work on the spring allowed them to work out that a tunnel dug by King Hezekiah in the eighth century BC to bring the water safely into the ancient city was used in a different way from that previously supposed by archaeologists. In addition, they uncovered a pair of massive towers dating from the eighteenth and seventeenth centuries BC

— demonstrating that the city was large and wealthy at that time^{2,3}.

The bullae came from a new salvage excavation near the spring, this one initiated by the need to unblock the sewerage line that dumps the Old City's waste in the Kidron riverbed. In 2005, the pair made another major find — the Pool of Siloam, mentioned in later Jewish sources and the New Testament.

This excavation ran up against the site's political and legal complexities. The excavation goes into the hill and under a mosque and a kindergarten, whose walls have cracked. The Palestinians say that the excavations were carried out without regard for their property and caused the damage.

Yigael Ben-Ari, district manager for the Israel Nature and National Parks Protection Authority's central region, says that the damage is unrelated and rejects the charge of indifference to the Palestinians and their property. "We agreed in advance to take care of any damage that the sewerage work and excavations would cause," he says. "An engineer we brought to examine the damage said there is no connection between our work and the cracks in the mosque. The work has turned up a find of worldwide importance." Ben-Ari says he and his staff have met with the Palestinian residents of Silwan and will help to repair the damage. "We can't run any site if the residents don't want us," he says.

Meanwhile, Greenberg and his colleagues maintain that the extent of the salvage excavations has been dictated in part by Elad's desire to create tourist attractions to present to the public. They argue that the site should be run by the parks authority or some other national body that is subject to public oversight and does not have a political or religious agenda.

In fact, Greenberg's group claims, such a transfer of authority was supposed to happen in 1998, when a group of 33 archaeologists petitioned Israel's Supreme Court to cancel the permit that allowed Elad to operate the site. The suit was withdrawn after the state attorney told the court that Elad's permit had been cancelled.

But in 2002, under legal circumstances that are still unclear, Elad resumed managing the park. This has spurred Greenberg and his group to prepare to renew the 1998 suit. According to Greenberg, the petition has been delayed for technical reasons but will be filed with the Supreme Court in the near future.

For his part, Reich says he disagrees with Elad's politics and belongs to the group of archaeologists who think that the evidence on the ground fails to support much of the Bible's narrative. But he has not joined Greenberg's efforts to remove Elad from involvement. The digs and the resulting tourism have provided employment for the Arab residents, he notes. Unlike many other archaeologists, Reich has not used student volunteers, and instead employed 20–30 local Palestinian residents — whose salaries were paid by Elad. "It's the residents' luck that they happen to live here," he says. "So they should live off the site as well."

The excavations could not be accomplished without the money and sponsors Elad brings

in, Reich maintains. He says that the group does not impose its politics or religion on him or other archaeologists. Furthermore, Elad has been willing to pay for the kind of unexciting, but crucially important, technical work for which it is difficult to find national and academic funding. For example, Elad has helped Reich to find a private funder to pay

a draftsman to copy the marks on the bullae, so that they can be analysed and compared.

Greenberg and his associates are still in the process of preparing their court petition. In the meantime, the conflict between science and politics at this most sensitive of archaeological sites has not affected either its interest to scholars or its attraction for tourists. Excavations continue in the City of David and the archaeological park is filled with Israeli and foreign visitors. And whatever the final decision on who runs the site, that is likely to continue.

Haim Watzman is a writer based in Jerusalem.



Rafi Greenberg wants to see different management at the City of David site.

1. Mazar, E. *Biblic. Archaeol. Rev.* **32**, 16–27, 70 (2006).
2. Reich, R. & Shukron, E. in *New Studies on Jerusalem, Proceedings of the Fourth Conference, Bar-Ilan University, Ramat-Gan, 1998* (ed. Baruch, E.) 5–16 (1998).
3. Reich, R. & Shukron, E. *Biblic. Archaeol. Rev.* **25**, 22–32, 72–73 (1999).



Work at the Africa Centre has influenced the WHO's infant feeding guidelines

A tale of two centres

Two institutes on opposite sides of South Africa are intent on tackling HIV. But they are separated by more than geographical distance, finds **Michael Cherry**.

Sandwiched between the green Hluhluwe-Umfolozi game reserve and the national motorway running north from Durban is a 435-square-kilometre dustbowl called Umkhanyakude. In the apartheid era it was designated a part of the Zulu homeland, and today each of the small plots is scattered with thatched huts and is home to a family.

Umkhanyakude has one of the highest levels of HIV prevalence in South Africa, peaking at 51% for women aged 25–29, compared with about 30% for pregnant women nationally¹. These data were collected by researchers at the Africa Centre for Health and Population Studies, a ten-year-old institute housed in an incongruously modern building in the centre of this impoverished region.

At the southwestern tip of the African continent, 2,000 kilometres away, lies fair-weathered and affluent Cape Town — another city suffering from the HIV epidemic but home to a rather different research centre. In Cape Town, HIV has spawned a higher incidence of tuberculosis (TB) than in any other city in the world. This disease is the subject of study by researchers at the five-year-old Institute of Infectious Disease and Molecular Medicine (IIDMM), part of the University of Cape Town. Last year, a team there sequenced DNA from the preserved organs of TB victims and

showed that the prevalence of a particularly virulent strain originating from southeast Asia has exploded since 1996. This strain is now causing 20% of TB deaths, perhaps because it acquires drug resistance particularly quickly.

Both the Africa Centre and the IIDMM are at the heart of South Africa's battle against HIV and both have received significant funding from Britain's Wellcome Trust medical charity. "Clearly there is some overlap between the two institutions," says Jimmy Whitworth, head of international programmes at Wellcome.

But the two also have their differences: the Africa Centre focuses on understanding the health of the local population and the best interventions to help this community, whereas researchers at the IIDMM are interested in understanding HIV on a molecular level and turning that knowledge into clinical practice. And despite the institutes' overlapping remit and the scale of the HIV crisis in South Africa, there is very little collaboration between them.

No one is prepared to say that this remote relationship is holding back research into the country's HIV epidemic. But it is "a cause for regret," says Robert Wilkinson, a senior fellow at the IIDMM who led the TB study, and it is an issue that he and others want to address.

"I don't think that the lack of collaboration is intentional," says IIDMM director Greg Hussey.

"The Africa Centre presents a huge opportunity to do meaningful research in an HIV-stricken area."

— Marie-Louise Newell

"The reality is that we tend to collaborate with overseas researchers because, with a few exceptions, they provide our funding."

The Africa Centre was founded by the Wellcome Trust in 1997 and is now part of the University of KwaZulu-Natal (UKZN). About 80% of the institute's running costs — currently US\$5 million — are covered by a grant from the trust. The centre faces a difficult balance between the demands of funders — who want high-quality publications in order to justify their support — and the needs and expectations of the poor rural community in which it is based.

Location, location, location

The Africa Centre's location has proved a major barrier to recruiting and retaining senior staff, particularly skilled South Africans who tend to be more interested in working in cities or abroad. Another problem in this regard is that the centre offers no security of tenure; like other Wellcome-funded institutes, all of the staff are on three- to five-year contracts — centre director Marie-Louise Newell is herself on a five-year secondment from University College London in the United Kingdom.

The IIDMM, by contrast, has proved to be a magnet for drawing South Africans abroad back to the country and for attracting foreign talent. The institute was established in 2002 as a virtual entity comprising researchers from departments in the University of Cape Town's science and medical faculties. Two years ago, a new building was completed to house the institute. Its annual budget is approaching US\$20



OBED ZILWA/AP; A. BUCKLAND



Clinical practice: the IIDMM is trialing a vaccine against South Africa's dominant HIV subtype.

million, about 40% of which is funded by the university and the rest of which comes from research agencies. The Wellcome Trust has so far supplied more than US\$12 million, mostly by funding senior fellowships. Although these are for only five years, the university has agreed to convert them into tenured appointments at their termination, subject to the fellowship recipients' good performance. The combination of the IIDMM's location, its secure and generous funding, world-class facilities and access to HIV patients all serve to attract good researchers.

The institute's mission is to translate molecular laboratory research into the clinic and it is having some of its greatest impact in African initiatives to develop vaccines against TB and HIV. Virologist Carolyn Williamson, now a principal investigator at the IIDMM, was central to the initial elucidation of the virus's diversity in South Africa. In 1999, her group showed that the country's epidemic is dominated by a strain of HIV known as subtype C, whereas in the developed world subtype B is dominant². This suggested that vaccines developed against foreign strains might not work well in South Africa.

The big picture

One of the Africa Centre's main achievements, meanwhile, are its HIV surveillance data. These represent one of the few precise estimates of HIV prevalence in South Africa, because they are collected by visiting and testing everyone in a household rather than, more typically, by testing pregnant women who attend antenatal clinics. In one study, the researchers compared their surveillance figures with those from an antenatal survey in the same region and confirmed that HIV prevalence was similar across the data sets in almost all age groups. But according to the surveillance figures, women aged 15–19 had an HIV prevalence that was only half that suggested

by the antenatal survey. This disparity has been attributed to the fact that some of the cohort are not yet sexually active³.

Another widely acclaimed accomplishment of the centre is a study on HIV transmission during breast-feeding. Previous studies estimated that breast-feeding is associated with a 10–20% probability of HIV transmission, so the World Health Organization (WHO) formerly recommended feeding infants with formula milk as a first choice, where it was safe to do so. But as in much of Africa, most women in Umkhanyakude lack access to clean water and cannot safely avoid breast-feeding, although many supplement it with formula or solid food. Previous studies had not distinguished between transmission risks from exclusive breast-feeding and this type of mixed feeding.

The centre's study showed that, in this rural setting, exclusive breast-feeding is a much better strategy. Babies who received breast milk and formula were nearly twice as likely to acquire HIV — and when mixed with solids, the risk was almost 11 times higher⁴. One reason for this is that infant formula and solid food can cause microscopic damage to cells lining a baby's immature gut, helping HIV to enter the body, says author Nigel Rollins of UKZN. The study helped prompt the WHO to refine its infant feeding guidelines earlier this year,

so that exclusive breast-feeding is the default option where feeding with infant formula alone is not possible or safe.

While the Africa Centre has focused on means of HIV transmission in the local community, Williamson's group at the IIDMM set to work developing vaccines against the dominant subtype C. The team surveyed HIV from newly infected patients and selected genes that best represented South African subtype-C strains⁵. The idea is that individuals immunized with vaccines containing genes from dominant local viruses are more likely to recognize and launch an immune response when infected by one of them. The IIDMM is now carrying out one arm of a clinical trial in Cape Town of an HIV vaccine that includes these genes, funded by the International AIDS Vaccine Initiative. The results of this trial should be available by the end of the year.

Building on this work, the HIV vaccine-development group at the institute, headed by Anna-Lise Williamson, recently developed two new, refined vaccines incorporating additional subtype-C genes and modifications to make them more potent⁶. These will go into clinical trials next year — the first HIV vaccines developed in Africa to enter trials.

Dual aspect

The Africa Centre presents a huge opportunity to do meaningful research in an HIV-stricken area, says Newell, and some researchers are hopeful that the relationship between the two institutes could yet blossom. Populations studied by the Africa Centre could prove an ideal testing ground for interventions developed at the IIDMM. "An obvious area of collaboration is the clinical testing at the Africa Centre of candidate vaccines that were developed at the IIDMM," says Lynn Morris of the National Institute for Communicable Diseases in Johannesburg. And, says Wilkinson, "it might well be possible to develop concrete plans for collaboration in the future."

As yet, there are no firm plans for such a partnership. But perhaps in time the 2,000 kilometres between Umkhanyakude and Cape Town will be bridged.

Michael Cherry is Nature's contributing correspondent in South Africa.

"It might well be possible to develop concrete plans for collaboration."
— Robert Wilkinson



1. Welz, T. *et al.* *AIDS* (in the press).
2. van Harmelen, J. H. *et al.* *AIDS Res. Hum. Retroviruses* **15**, 395–398 (1999).
3. Rice, B. *et al.* The XVI International AIDS conference, Toronto — Track C15 — HIV/AIDS surveillance TUPE0312 (2006).
4. Coovadia, H. M. *et al.* *Lancet* **369**, 1107–1116 (2007).
5. Williamson, C. *et al.* *AIDS Res. Hum. Retroviruses* **19**, 133–144 (2003).
6. Burgers, W. A. *et al.* *J. Gen. Virol.* **87**, 399–410 (2006).

See Editorial, page 1.

Long shadow of Linnaeus's human taxonomy

SIR — Your 15 March issue honouring Carl Linnaeus brings to mind what is probably his most significant contribution to modern life: the idea that groups of people can be regarded as naturally distinct taxonomic entities, or subspecies, in the same fashion as species, genera and higher categories.

In the first edition of *Systema Naturae*, published in 1735, before formalizing binominal species nomenclature, Linnaeus presented humans as sorting naturally into whitish Europeans (*Homo Europaeus albens*), reddish Americans, dark Asians and blackish Africans. By the 10th edition, in 1758, these had become subspecies, colour-coded as red Americans, white Europeans, yellow Asians and black Africans. In addition, Linnaeus separately listed wild children (*Homo sapiens ferox*) and a non-geographical grab-bag category, *Homo sapiens monstrosus*.

The idea that humans can be understood as constituting natural taxonomic units has bedevilled anthropology ever since. In the eighteenth century, both Johann Friedrich Blumenbach and Georges-Louis Leclerc, Comte de Buffon, recognized that the principal empirical patterns of human diversity are geographically gradual, which frustrates the project of human taxonomy. We would now say that pattern is 'clinal'¹.

Further, as anthropology matured, it clarified the fact that human groups principally differentiate themselves culturally — that is, by language, dress, principal deities, taboos and the like. The strongest antagonists are not the most biologically different, but simply the worst neighbours. Consequently, perceptions of group difference are local, political and ephemeral; but are nevertheless still commonly regarded as natural — witness the racialization of categories such as 'Hispanic' and 'Middle Eastern' in the United States, and the newsworthy discovery in the United Kingdom that the Irish and the Scots are not so different after all².

Genetics has been multivocal on the subject. On the one hand, it has emphasized the extensive polymorphism in the human gene pool³, showing that there are all kinds of people everywhere — as fieldworkers had long known, but without quantitative support. On the other, it has focused on the small component of genetic variation that differs the most in the most widely separated peoples, and commonly interpreted the results in racial terms⁴.

It has taken two and a half centuries to shed Linnaeus's fallacy that the human species comes taxonomically organized into a few large, natural groups that are fairly homogeneous and fairly distinct from one another. We have come to understand,

rather, that the predominant patterns of human variation are cultural, polymorphic, clinal and local.

This does not mean that everyone is the same, or that there is no biogeographic differentiation within our species. It means just that the effort to treat our own species taxonomically has considerably more social and symbolic than biological meaning.

Jonathan Marks

Department of Sociology and Anthropology, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA

1. Huxley, J. S. *Nature* **142**, 219–220 (1938).
2. Oppenheimer, S. *The Origins of the British: A Genetic Detective Story* (Carroll & Graf, 2006).
3. Lewontin, R. C. *Evol. Biol.* **6**, 381–398 (1972).
4. Bamshad, M. J. et al. *Am. J. Hum. Genet.* **72**, 578–589 (2003).

Brain drain: gains all round when it goes both ways

SIR — In your Editorial "In praise of the 'brain drain'" (*Nature* **446**, 231; 2007), you ask how UK science flourishes despite the continual brain drain to California and elsewhere. One answer is the compensating in-drain from the Commonwealth, the rest of Europe and elsewhere.

Many universities in mainland Europe are dysfunctional in terms of the career ladder for young academics, influences in appointment, departmental management structure, research funding or senior management structure.

Two comments from people in my field: "I could never get an academic position in my country because my PhD supervisor is not good at academic politics"; and "In my [different] country, research funding is spread equally among all groups without regard to quality, and those doing outstanding work cannot get any more." Both these individuals have good positions in the United Kingdom and one was promoted rapidly to a personal chair at Cambridge University. There are many other examples.

Instead of complaining about the brain drain out, we should be encouraging the brain drain in. All PhD research studentships could be open equally to anyone in the world. Even those who then go back to their home country make a contribution in addition to the work they have done here: in a few years' time they start sending us their best output as PhD students or young postdocs, and the cycle repeats itself, with some of the new crop staying on.

I agree with your Editorial that the situation is good for the home countries, in the sense that there is a pool of people at the world forefront, available to be enticed back for senior appointments, and there is a pipeline for training the new students. If all the New Zealanders with good scientific

jobs around the world tried to return to New Zealand, the country would burst!

Volker Heine

Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK

Brain drain: poor countries lose most and benefit least

SIR — Your Editorial "In praise of the 'brain drain'" (*Nature* **446**, 231; 2007) is, in my opinion, misleading in its representation of the issue and in its attempt to justify phenomena that are debilitating for the education and training of professionals in the developing world.

The effects of these phenomena on countries differ, depending on the extent of a country's development. Developed countries, on the whole, have large numbers of scientists and healthcare and other professionals in their populations, whereas developing countries may have just a handful. A major obligation of any government to its population is to pursue and implement policies that increase numbers of these key professionals to a desirably stable level, or — where they are already approaching stability — to maintain them at those levels. This responsibility on governments is independent of international opinion and is the reason why South Africa may decide to penalize individuals who leave the country after having been trained at the state's expense, or rebuke companies that facilitate the mass emigration of its professionals.

The gains from money sent back home, or from some expatriates returning to their native countries much later in their careers, may be of some benefit to those countries. But it is difficult to imagine how this could be more useful than doctors, nurses, teachers or lecturers staying in places where such people are seriously lacking.

You mention the correlation of higher emigration rates with better public healthcare systems, but a correlation is not a causal link. What else would we expect when the countries that 'drain the brains' have the power to pick and choose? When migration occurs between countries that have no large disparity in their development levels, the exchange is more likely to be mutually beneficial.

Given all this, I believe that there is nothing to praise about the brain drain when it occurs en masse from the developing countries into richer, more developed ones with dramatically more power.

Andrew Isaac Meso

Royal Holloway University of London, Egham, Surrey TW20 0EX, UK

Contributions to Correspondence may be submitted to correspondence@nature.com.

SPRING BOOKS

A strange sense of self

Am I a mirage?

I Am a Strange Loop

by Douglas R. Hofstadter

Basic Books: 2007. 384 pp. \$26.95, £14.99

Susan Blackmore

Who is this 'I'? Is it the author? Is it the reader? Could it be the sentence itself (would that work)? Readers of Douglas Hofstadter's best-seller *Gödel, Escher, Bach* (Basic Books, 1979) will be familiar with such twisty questions, and will probably delight in the pain, mental squirming and occasional wondrous resolution that they provoke. But whereas *Gödel, Escher, Bach* conjured myriad mysteries, *I Am a Strange Loop* tends rather to explain them, and the result is less magical and even slightly awkward.

At one level (and this is a deeply multilayered work) the book pulls off some remarkable achievements. For example, in a matter of 40 readable, and even enjoyable, pages, Hofstadter manages to explain Kurt Gödel's incompleteness theorem in a way I have never seen attempted before. We are taken from "the gloomy, austere, supposedly paradox-proof castle" of Alfred North Whitehead and Bertrand Russell's great work *Principia Mathematica* (containing the formal system that Hofstadter calls 'PM'), through ways of mapping numbers onto theorems and hunting for patterns among squares, primes and 'prim's' (Hofstadter's name for theorems that are provable in PM), to sentences that talk about themselves and so, finally, to make sense of a concise English translation of Gödel's formula (the paradigmatic strange loop), "I am not provable in PM." Wow! I really felt I learnt a lot.

Hofstadter speculates that Russell never saw the second level of meaning (the effect of mapping numbers onto theorems) in his great work, like a dog that sees a television screen as a mass of changing pixels, or a child who sees the people on the screen but fails to grasp the romantic plot. And then he whisks us away to tangle with ever more layers of paradox and wonderfully mind-wrenching questions. What is the nature of mathematical truth? What is the nature of meaning? Could a machine be confused? Could it know it was confused? Could it believe that its unquestioned belief in the reality of its own 'I' is a necessary illusion?

Along the way, Hofstadter talks about himself, but his pacy mix of stories, metaphors, questions and explanations is sometimes spoiled by what seems like a lack of confidence.

Instead of just blasting us with his rush of original ideas, Hofstadter apologizes for a "corny pun", a "hopefully amusing example" or just for telling personal stories at all. Yet these stories are delightful. My favourite is his first encounter with something that, he says, "runs in our human grain": the irrational fear of loops. When little Duggie went with his parents to buy a video camera, one of the cameras in the store was plugged into a television screen. So he pointed the camera at his father, then at himself and then... about to point it at the screen itself, he stopped. He remembers with shame that he was hesitant to close the loop. So he timidly asked the salesperson whether he might and was told: "No, no — you'll break the camera!"

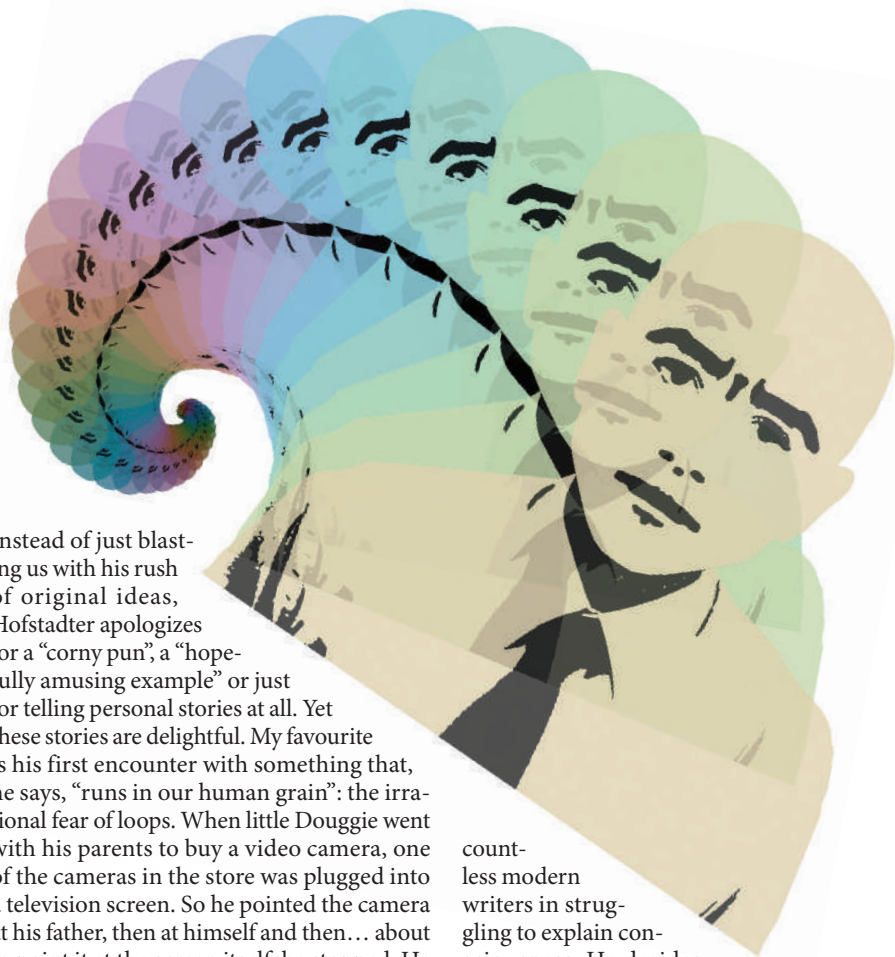
He then convinced himself that this couldn't possibly be true, and went home to experiment with video feedback, finite and infinite regresses, corridors with curves or corners, and a completely unexpected pattern with precisely 13 (a prime number, of course) spokes.

A sadder event, the discovery of his baby sister's brain damage, began his fascination with the physical basis of consciousness. At the age of 12 or so, it dawned on him that consciousness is a peculiar kind of mirage that perceives itself and yet doesn't believe it's perceiving a mirage. This insight leads directly to the stated aim of this book: to try to pinpoint that "special kind of subtle pattern" that underlies, or gives rise to the 'soul', the 'I', 'having a light on inside' or 'being conscious'.

This is a grand aim, and Hofstadter joins

countless modern writers in struggling to explain consciousness. He derides zombies and qualia, has harsh words for philosophers David Chalmers and John Searle, and skilfully sweeps away all sorts of nonsense, from old-fashioned kinds of dualism to the more prevalent belief that consciousness is still something 'extra' — an *élan mental*. Instead, he argues that the self is a strange loop that automatically arises in a machine with a sufficiently sophisticated repertoire of categories. It is a myth, a mirage, like a satellite to your brain whose resident strange loop decides that 'here' is wherever that brain happens to be. And, he claims, once you have explained the self, you have explained consciousness.

Herein lies the source of my dissatisfaction. The idea of the self as a strange loop makes sense of moments of self-awareness and of baffled self-inquiry — but what about the rest of the time? The theory seems to imply that mostly we are not conscious at all, which may well be right, but Hofstadter does not discuss this. Then there are those profound moments



ILLUSTRATIONS BY JOE MAGEE

of utter stillness or absorbed flow when the self is in abeyance. People describe these as being clearer than ordinary consciousness, but this cannot be explained if self and consciousness are as closely linked as Hofstadter claims. He also argues that the self loop is indispensable; this might be challenged by those who have attempted, or even managed, to let go of the illusion of self. He quotes a Zen koan that seems beautifully to point the way out of strange loops and into awareness beyond self, but he dismisses it as “just a bunch of non-sequiturs”.

Hofstadter realizes that people will be dissatisfied, and provides a light-hearted debate

between two numbered strange loops: himself and his opposing sceptic. I keep wondering whether I’m no better than sceptic SL #264, who really just doesn’t get it, although he (it?) comes up with some classic objections.

I keep looking over at my black cat on the window sill, with her ragged fur brightly lit by the morning sun. Does it help to say that the experimenter of this vivid visual experience is a strange loop? Whatever the answer, this strange loop is enjoying the question.

Susan Blackmore is a freelance writer, lecturer and broadcaster, and a visiting lecturer at the University of the West of England, Bristol, UK.

Dawkins and many others. Angier’s distinction lies in her exhilarating use of language. Unlike Bill Bryson’s *A Short History of Nearly Everything* (Doubleday, 2003), *The Canon* does not rely on personalities to brighten up the prose, even when exploring traditionally difficult areas of physics and chemistry. Instead, anecdotes are well chosen, humanizing without patronizing the scientists involved. Angier has no need to name-drop; her writing style holds the reader’s attention. (I can vouch for this, having almost missed my stop while reading her book on the train.) For once the blurb — “playful, passionate” — is spot on: this is an astonishingly literary science book, much better written than most. Out goes pedestrian prose; in come references to every cultural form from the scriptures to movies, delicate allusions to writers from Homer and Andrew Marvell to Sigmund Freud, such words as ‘accoutred’, ‘trocar’ and ‘miasmatic’, teasing alliterations, the occasional sharp political comment and some truly excruciating puns.

This linguistic fecundity can at times be overwhelming, especially for non-American readers, who may find some of the references baffling (Ty Cobb? Bialies? You’d best keep a search engine handy). A riff on chemical bonds by analogy to their superspy namesake James was unconvincing and somewhat distracting — if Sean Connery is covalent and Roger Moore ionic, where does that leave Daniel Craig? — but this is one of the book’s few weaker moments. Elsewhere there are glories, as when Angier remarks of sexual selection in the peacock: “If you survive long enough to breed, and if you score handsomely, even orgiastically, in a single spring spree, who cares if you’re a feather duster come summer?” The peacock’s tail is standard fare in evolutionary biology textbooks, but few descriptions linger in the mind as enjoyably as Angier’s.

Readers who like their texts spartan will loathe this one, which rarely uses one metaphor, simile, adjective or subclause where two, three or four can be squeezed in. Purists who object to anthropomorphisms should also take note: *The Canon* overflows with loyal water droplets, preening cells, anxious and fidgety electrons, and the like. In my judgement, however, the benefits — if only in counteracting the still-prevalent ‘two cultures’ stereotype of science as the preserve of barely literate philistines — make the purple passages worthwhile. Those

James Bond with a feather duster

The Canon: A Whirligig Tour of the Beautiful Basics of Science

by Natalie Angier

Houghton Mifflin: 2007. 320 pp. \$27

Kathleen Taylor

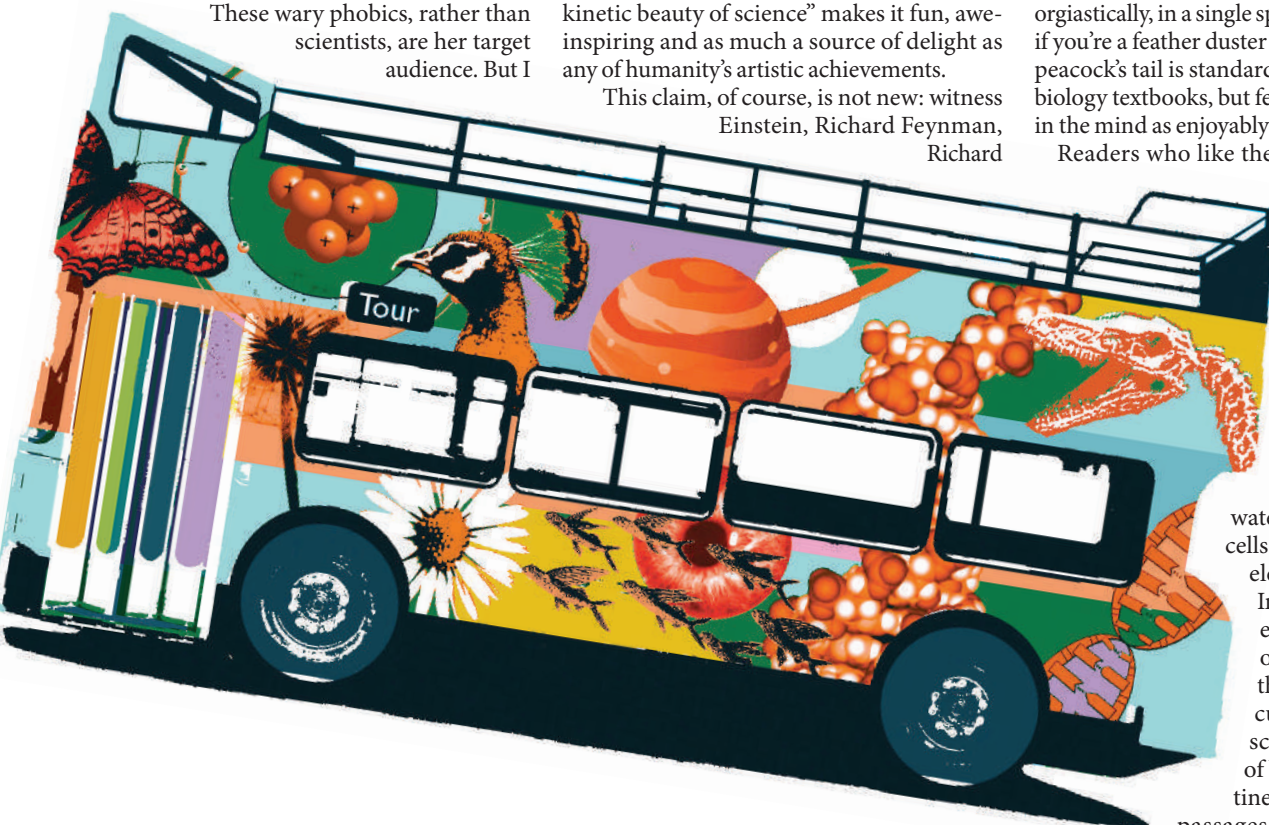
Natalie Angier’s book *The Canon*, like many before it, sets out to persuade the public that learning about science can be enjoyable. It focuses on the ‘hard’ sciences, which here means physics, chemistry, evolutionary and molecular biology, geology and astronomy, together with introductory chapters on thinking scientifically, probability and issues of scale and measurement. Angier proposes that what scientists do is worth a look even for people traumatized by school science lessons.

These wary phobics, rather than scientists, are her target audience. But I

would also recommend *The Canon* to professionals, and to the already interested public (a sizeable constituency, as not all school science teaching is bad), because this is a remarkable and delightful book.

Angier, an accomplished, Pulitzer Prize-winning science journalist, has clearly thought carefully about the ‘why bother?’ challenge to science communication. She notes but does not depend on the common arguments that the importance of science makes avoiding it unjustifiable; that future national prowess requires more scientists; or that a scientifically informed public may at last learn reason and decide to put astrologers and lotteries out of business. Instead, she eschews “civic need” for “neural greed”, aiming to demonstrate that “the kinetic beauty of science” makes it fun, awe-inspiring and as much a source of delight as any of humanity’s artistic achievements.

This claim, of course, is not new: witness Einstein, Richard Feynman, Richard



who find the style off-putting should persevere, because Angier's gift for metaphor lights up the dustiest corners. Her explanation of how electromagnetic radiation is produced is superbly easy to visualize; she is lucid on evolution, and on intelligent design; and her chapter on molecular biology is an exemplary introduction. Science teachers should find numerous useful resources here. Instead of relying on geeks-and-gimmicks clichés — eccentric geniuses, bitter feuds and zany facts — Angier's word-painting allows the scientific material to speak for itself in some depth. Consequently, we get a real sense of science as an immense collective endeavour, comprising both established knowledge and works-in-progress, done but not entirely dominated by personalities.

This is not a cutting-edge specialist text, so its contents are likely to be familiar to *Nature*

readers. Much is necessarily omitted; the chapter on astronomy in particular feels disappointingly slight. Nevertheless, as an introductory guide, *The Canon* sets the standard for science writing and deserves at least to be shortlisted for the Royal Society Prize for Science Books. Its style may seem densely, even formidably, allusive at times, but Angier's gift for accessible explanation is outstanding. If any book can help the public learn to love science, this is it. ■

Kathleen Taylor is in the Department of Physiology, Anatomy and Genetics, Oxford University, Parks Road, Oxford OX1 3PT, UK.

The dark heart of the bomb

Plutonium: A History of the World's Most Dangerous Element

by Jeremy Bernstein

Joseph Henry Press: 2007. 258 pp.

£16.99, \$27.95

John S. Rigden

Plutonium has either a celebrated or a tragic history, depending on your point of view. It was the core of the weapon that destroyed much of Nagasaki on 9 August 1945, and has only military uses. For those who find security standing behind a stockpile of plutonium bombs, the element is a reason to celebrate. By contrast, for those who regard the bombing of Nagasaki as a needless repetition of the Hiroshima catastrophe, plutonium is a symbol of the US–Soviet arms race that dominated the second half of the twentieth century. It now signifies the rank and status of a nation's military prowess.

In his book *Plutonium*, Jeremy Bernstein acknowledges that everything connected with the element is complicated, and that includes plutonium itself and its history. Its discovery in 1941 by Glenn Seaborg and Arthur Wahl is part of a much bigger story in which each part becomes a story in itself.

Plutonium does not occur in earthen deposits, for example; it is produced instead by the radioactive decay of uranium by way of neptunium, and it is with uranium that the book begins. Then there is the story of the periodic table and the problems associated with fitting the elements into their proper places — especially the lanthanides (the elements of atomic number 58 to 71 that follow lanthanum in the periodic table) and the actinides (elements 90 to 103 following actinium). There is the story of radioactivity (and the connected story of the discovery of X-rays) and of Enrico

Fermi bombarding uranium nuclei with slow neutrons. Add to these the story of fission, with various elements and isotopes complicating the plot. Los Alamos and the development of atomic bombs are also a central part of the plutonium story. Finally, there are the complications arising from the element plutonium itself that must be understood and the associated problems solved. Melding these many parts into a short book represents a daunting challenge, which Bernstein confronts head on.

One of the benefits of this multifaceted approach is the opportunity it gives the author to educate readers by means of historical information and thumbnail sketches of interesting people. In his 1903 Nobel address, for example, Henri Becquerel, who discovered radioactivity, suggested that the energy associated with radioactivity may involve the modification of atoms in the radioactive material. Two years later, Einstein showed that there was a loss of mass, which becomes energy according to his famous equation $E=mc^2$. In 1934, Ida Noddack correctly criticized Fermi, suggesting that in his neutron-bombardment experiments he had actually discovered nuclear fission. Fermi's Nobel speech in 1938 was wrong on this point because he assumed he had discovered transuranic elements. When the Nobel Prize was awarded for the discovery of fission, the Nobel committee made so many erroneous assumptions about who did what, and when, that



Lise Meitner was wrongly denied a share of the prize.

The tale of Fritz Houtermans is particularly interesting and not well known. Houtermans wrote a report in 1941 in which he considered the absorption of a neutron by uranium-238 and concluded that it would lead to plutonium via neptunium. He further concluded that plutonium would be fissionable. Perhaps generalizing from his own insights, he twice sent messages (from his native Germany) to the Allies that Germany was “on the track” to making plutonium. It would be interesting to know why he did this, but Bernstein says only that he wanted to “warn the Allies”. In any event, Houtermans was wrong: the Germans were not close to making plutonium.

In early 1943, the Los Alamos laboratory — the home of the Manhattan Project — began to take shape. By the summer of 1944, plutonium started arriving there. The element's

idiosyncrasies and complexities soon became apparent. William Zachariasen discovered that plutonium had six different crystal structures, or allotropes, which he labelled α , β , γ , δ , δ' and ϵ . One of these allotropes had to be formed into a metal suitable for a bomb, which meant being stable and free of isotopes that would interfere with a chain reaction. The metallurgist Cyril Stanley Smith had the good fortune and acute intuition (there were no data) to select gallium to form an alloy with the δ allotrope of plutonium to produce the needed stability.

It was still unclear whether the δ allotrope would revert to the α allotrope before explosion. And a way of bringing the two subcritical pieces of plutonium together to form the critical mass — and initiate the chain reaction that would lead to a nuclear explosion — had to be developed from scratch, as the gun trigger used for the uranium bomb that was dropped on Hiroshima was not suitable. Plutonium, then, presented challenges at every turn. As Bernstein suggests, it may have been only the fear of what the Germans were doing that kept

the physicists working long into the night.

This book will make demands of readers. There are many things to hold in the mind as Bernstein repeatedly moves away from the main thrust of the book to develop one of these side stories, which enrich the story of plutonium but are also sometimes a distraction. But Bernstein's writing ability smooths the way and makes this a successful book. ■

John S. Rigden is in the Department of Physics, Washington University, St Louis, Missouri 63130, USA.

Trouble in paradise

Coral: A Pessimist in Paradise

by Steve Jones

Little, Brown: 2007. 256 pp. £15.99

Daniel Pauly

If I had first seen *Coral* by Steve Jones in a bookshop, rather than receiving a review copy, I would have bought it. I would have been attracted by its superb cover, whose eerie blue serves as a glorious background for a swimming red snapper. And attempting to casually browse through the text, I would have been slowly ensnared by the loops of its fascinating literary, historic and scientific digressions.

Any book with the word 'pessimist' in its title must have a sound basis. Here it rests on Charles Darwin's solid shoulders — or more precisely, on his first scientific book, from 1842, *The Structure and Distribution of Coral Reefs*, in which he presented a hypothesis that solved the riddle of how coral reefs grow, where they grow, and why. Jones, in his first chapter, explains how Darwin came to his hypothesis, how it shaped all subsequent research on coral reefs, and how drilling into Pacific atolls, conducted in support of nuclear bomb tests, ultimately confirmed it.

Darwin's book

relied on the simple but profound idea that 'lowly' organisms, here coral polyps, pursuing their own tiny goals, through their sheer numbers and over the immensity of time, could play major roles on the geological stage. This is also a theme in his 1859 book *The Origin of Species*, whose detractors could not fathom the transformative power of small, between-generation changes occurring over eons. This simple idea was again the theme of his 1881 book on the slow, subterranean work of earthworms, *The Formation of Vegetable Mould Through the Action of Worms*, to which he devoted his final years.

Genomics has given us a powerful tool to study the phylogenetic history and affinities of these tiny agents of change. In his second chapter, Jones uses genomics and the hydra (a non-colonial polyp related to corals) to introduce the notion that the cells of hydra cooperate, just like those of people.

They do this, says Jones, because they have learnt from the mutually beneficial relationships of

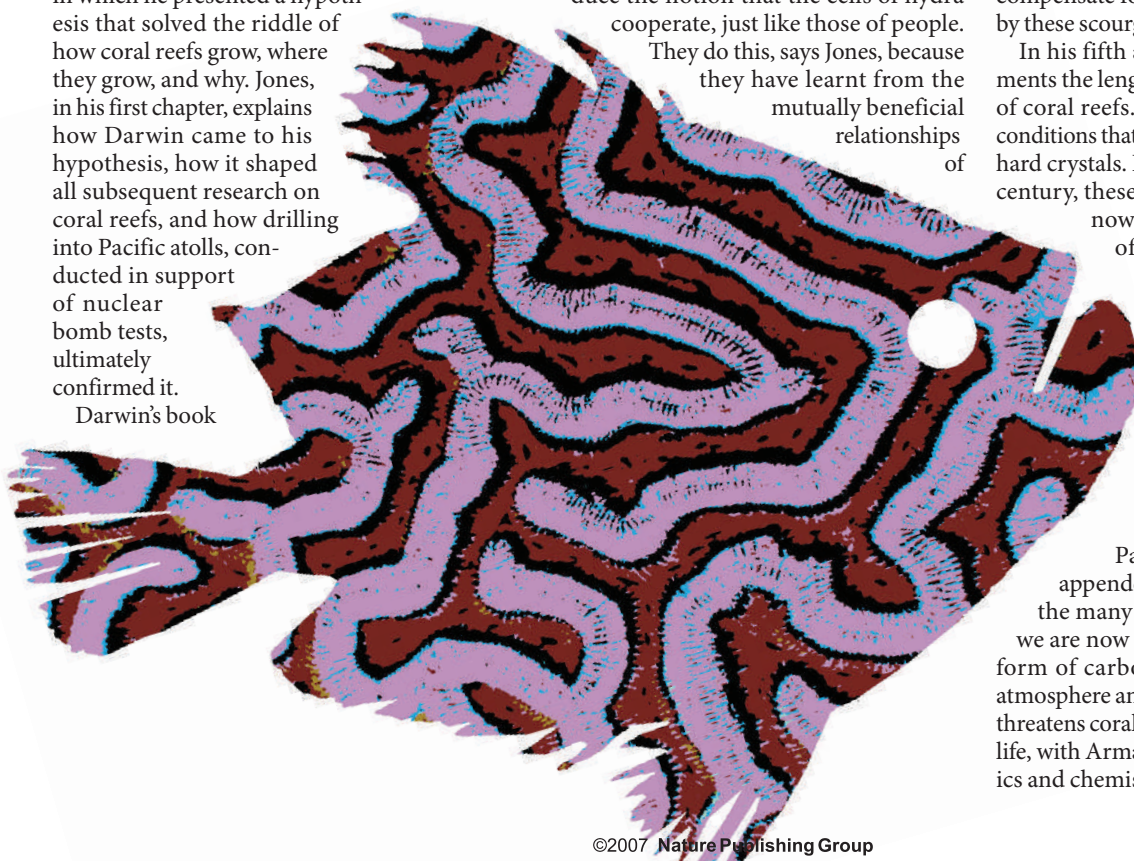
their organelles, many of which are descendants of formerly independent bacteria-like organisms.

He explores this idea further in the third chapter, which is devoted to what appears, in coral reefs and other ecosystems elsewhere, to be disinterested cooperation between species. But it isn't, notwithstanding the benevolent anarchist Prince Kropotkin, who gets a loop of several pages. Rather, barely masked warfare prevails, interrupted by tenuous and short truces, revoked when conditions change. Altruism seems to be limited to humans, and one of the biggest tasks we face is to expand our altruistic acts from our circle of relatives, friends and compatriots to the whole of humanity.

Jones then disposes, in his fourth chapter, of the tenacious Western myth of South Pacific coral islands as 'paradise'. Life was too precarious for that, particularly after the first contact with Europeans, who brought previously unknown diseases, some sexually transmitted. The abolition of cannibalism did not compensate for the population losses caused by these scourges.

In his fifth and final chapter, Jones documents the lengthy and rapacious exploitation of coral reefs. He starts with the geological conditions that cause carbon to form extremely hard crystals. In the middle of the nineteenth century, these conditions in parts of what is now India enabled the Maharajah of Hyderabad and his court to trade diamonds, via the East India Company, for jewellery carved from calcium carbonate from Mediterranean corals. Now the East India Company is no more, and these precious corals are mostly gone too.

Jones calls the book's epilogue, entitled 'A Pessimist in Paradise', an 'envoi', as if it were appended to a poem. He uses it to pull the many strands of this book into one: we are now stuck with trash carbon in the form of carbon dioxide that gums up our atmosphere and, as carbonic acid in sea water, threatens coral reefs, and indeed much marine life, with Armageddon. He explains the physics and chemistry involved with much verve,



and more looping (Captain Cook, Australian cockatoos, the Permian extinction, the prospect of 9 billion humans, the Irish Republican Army, California's abalone, Newton, Funafuti Atoll in Tuvalu...).

Finally, he explains his pessimism: "The world of coral gives more reason for despondency than for hope. Local conservation can do

little in the face of global change. The future of the reefs is bleak indeed. Their end presages a catastrophe that will spread far beyond their bounds — and remind us that we too are far from safe."

Daniel Pauly is director of the Fisheries Centre, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

Unearthing gender issues

The Invisible Sex: Uncovering The True Roles Of Woman In Prehistory

by J. M. Adovasio, Olga Soffer & Jake Page
HarperCollins: 2007. 320 pp. \$26.95

Pat Shipman

The Invisible Sex is a refreshing book that opens with a crucial reminder: "science is not truth; it is, instead, a method for diminishing ignorance." The authors — two well-known anthropologists (Jim Adovasio and Olga Soffer) and a science writer (Jake Page) — set out to diminish readers' ignorance about the human past, using a breezy, colloquial style that only occasionally irritates.

Their main point is that the male-dominated science of anthropology has usually chosen to interpret the evidence of the human past by basing it on male-dominated stereotypes. The authors succeed admirably in heightening the readers' awareness of such practices and in countering these stereotypical presentations with imaginative and equally defensible reinterpretations of particular sites or bodies of evidence. The end result helps to flesh out a more plausible female role in prehistory than has been offered previously. In many ways, this book is a much-needed antidote to the past hundred years of popular and scientific writing on prehistoric human life, and avoids the clichéd pitfall of veering too far into a hyper-feminist view.

The authors make many palpable hits. For example, they remark on the assumption that Lucy, the first largely complete *Australopithecus* to be found, was identified as female because the bones were small, not because they were diagnostically female. Similarly, they deconstruct the beautiful diorama at the American Museum of Natural History in New York of two australopithecines walking across the Laetoli plains, which are covered in damp volcanic ash, about 3.3 million years ago. A male and female walk together companionably, his arm around her shoulders. The female's head is turned,

giving the impression that she is looking at the viewer, while the male remains focused on what lies ahead. The subtext of this depiction emphasizes the intimacy of their closely spaced footprints with the familiarly possessive/romantic gesture of a male arm around the female's shoulders. What would fit the evidence just as well — and echo many more observations of primate social behaviour — would be that the footprints were made not by a 'couple', but by a female and her juvenile offspring. Why choose one over the other?

Another target for scepticism is the supposed dominance of a hunting lifestyle among hominins. This idea ranged from Raymond Dart's lurid osteodontokeratic hypothesis — that bones, teeth and horns were used with minimal alteration by early hominins as tools for slaughtering animals and possibly each other — through the presentation of early, tool-making *Homo ergaster* in East Africa as a hunter rather than a scavenger. This 'man the hunter' stereotype lingers in images of the mighty, mammoth-slaughtering Palaeo-Indians in North America. And yet, the authors

argue, the mere existence of tools does not prove that hunting was important, much less that it was the mainstay of hominin survival or a predominantly male activity.

Such biases of the prehistoric record are common, especially with regard to the oldest sites. The authors aver that in recent dry cave sites, fibre artefacts outnumber stone ones by a factor of 20 to 1. In several other situations, fibre and wood artefacts have been found to account for 95% of all artefacts recovered. That amounts to a tremendous amount of information not available to archaeologists in most parts of the world.

These are stunning observations that remind anthropologists that what we see is a tiny fraction of what might once have been present, not only in terms of individual animals but also in terms of artefacts. Many of those 'lost' artefacts may have been essential aids to gathering, capturing small (not heroically large) animals, or modifying the world (building nests or brush shelters) in ways that do not involve obtaining food.

The authors also review the fascinating discovery by Adovasio and Soffer of fibre impressions on the clay fragments at Dolni Vestonice I in the Czech Republic. These attest to the existence of eight different weaving techniques, sewing, net-making and basketry, providing a startling new glimpse of life 26,000 years ago. They suggest that fibre arts had been a well-developed industry for some time before the formation of that site.

Making things out of fibre is not the sole prerogative of either sex in ethnographic accounts, the authors point out. But throughout the tribal world today, women make most of the basketry. The making of ceramics items, especially pottery, is chiefly the province of women. So, they claim, it is safe to assume that most, if not all, of the ceramics, weaving, basketry and clothing was made by women in the years that Dolni Vestonice and the other Moravian sites were inhabited.

This is an astonishing leap of faith for those who have advocated a greater appreciation for the variability and malleability of gender roles elsewhere in the book. The interpretations offered by the authors are no more convincing than the standard ones, primarily because their interpretations are based on ethnographic and behavioural analogies that are different from, but not demonstrably sounder than, those they criticize.

Fortunately, they never grapple with the central and most difficult questions of all. For example, when is it justifiable to draw on behavioural analogies from modern humans to interpret the past? When ought we to rely on behaviours of non-human primates or other mammals instead? And how are we to evaluate



the relative probability of past behaviours when modern behaviours are so variable? Having pointedly raised these issues, the book disappoints because it does not offer any principles for doing a better job of reading the past.

The book is also flawed by appallingly poor attention to fact-checking. The authors misplace Olduvai Gorge and Laetoli in Kenya (they are in Tanzania); they call Mary and Louis Leakey palaeontologists instead of archaeologists; they cite *Homo* as the first of "what could at the time be called the hominid [hominin] line" (a grossly inaccurate remark

because australopithecines are also hominins or hominids); and they mistakenly state that the carbon isotopes that distinguish grass-eaters from tree-eaters are ^{14}C and ^{13}C , rather than ^{13}C and ^{12}C .

Unforgivably, there are no notes in the book to identify the sources of either direct quotations or specific points of information. These glaring defects undercut the credibility of this otherwise intelligent and provocative work. ■ Pat Shipman is adjunct professor of anthropology, Pennsylvania State University, 315 Carpenter Building, University Park, Pennsylvania 16802, USA.

to view as a primary task the maintenance of funding for the modelling juggernaut." Have the authors lost their nerve when discussing the politically sensitive issue of climate change? Even so, their strong views on modelling in climate science are difficult to miss.

Despite these quibbles, this is a valuable book for the very reason that it causes dissonance. Using well-documented cases of policy failure, the authors have identified a critical challenge confronting the modern scientific enterprise: our ability to produce model-based predictions seems to have outpaced our ability to use such tools wisely in decision-making. The results are seen in bad policies and bad science.

So what to do? The authors' plea for a world without models is unrealistic, and not simply because scientists will continue to produce them. A story related by the Nobel prizewinning economist Kenneth Arrow explains why. As a weather forecaster in the Second World

War, Arrow and his colleagues were told that

When the numbers don't add up

Useless Arithmetic: Why Environmental Scientists Can't Predict the Future

by Orrin Pilkey & Linda Pilkey-Jarvis

Columbia University Press: 2007. 230 pp. \$29.50, £19

Roger Pielke Jr

The central thesis of *Useless Arithmetic*, by the father-and-daughter team of Orrin Pilkey and Linda Pilkey-Jarvis, is "the virtual impossibility of accurate quantitative modelling to predict the outcome of natural processes on the Earth's surface". This is sure to cause cognitive dissonance among many readers — it simply does not seem to accord with our lived experience.

As I write this review, I'm sitting on an aircraft safely crossing the United States. The plane was created with quantitative aeronautical engineering design models, its flight path dictated by quantitative routing models, and the snowy weather I experienced at take-off was predicted by quantitative weather-forecasting models. Such experiences in successfully predicting and managing natural processes would seem to indicate that without mathematical models our twenty-first-century lives would simply be impossible. What could have influenced the authors to make claims so strongly contradicted by experience?

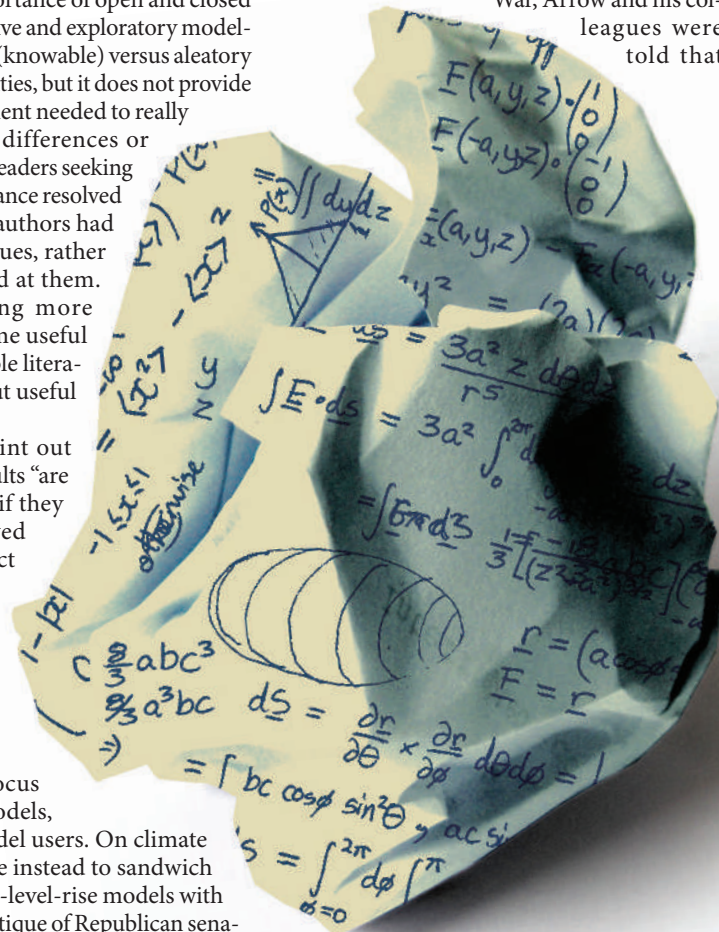
The authors answer this question unambiguously and definitively by discussing some two dozen instances (seven of which are examined in detail) in which decision-making, and in some cases science, based on mathematical modelling has gone awry, leading to undesired societal and environmental outcomes.

The difference between my flight and failed predictions and mismanagement of fisheries, say, is an important part of the book's argument. In engineering design, flight-path routing and weather prediction, quantitative models are produced and used in a different way from those used by policy-makers for fisheries management, sea-level rise, beach nourishment and the disposal of nuclear waste. However, many of the nuances and complexities involved in understanding these differences are hard to discern in the high-level,

non-technical overview provided in *Useless Arithmetic*. For example, the book hints at the contextual importance of open and closed systems, consolidative and exploratory modelling, and epistemic (knowable) versus aleatory (random) uncertainties, but it does not provide the in-depth treatment needed to really understand these differences or their significance. Readers seeking to have their dissonance resolved may wish that the authors had explained these issues, rather than merely hinted at them. But those wanting more depth will find some useful pointers to accessible literature in a concise but useful bibliography.

The authors point out that modelling results "are easier to live with if they follow preconceived or politically correct notions", and the chapter on sea-level rise related to human-caused global warming seems to bear this out. In most chapters, the authors focus on a critique of models, modellers and model users. On climate change they choose instead to sandwich their critique of sea-level-rise models with an even stronger critique of Republican senator James Inhofe and author Michael Crichton, both of whom have strongly taken issue with the science of climate modelling and action on climate change. Yet the following authors' comments would have been equally at home in one of Inhofe's speeches or in Crichton's sceptical novel *State of Fear* (HarperCollins, 2004): "the juggernaut, the large industry that has risen to answer questions about global climate change, global warming, sea level rise, and all their ramifications, has unstoppable momentum... leaders in global change studies tend

their commanding officer needed a long-term forecast. The forecasters knew from experience that such forecasts had little scientific basis, and related this up the chain of command. The reply that came back was this: no matter, the general needs the forecast for planning purposes. Quantitative predictions fulfil important political and social roles, regardless of their quality, accuracy or appropriateness, and will continue to be demanded by



decision-makers and produced by scientists.

If quantitative models are here to stay, an important question to ask is how we can improve the way we create and apply models in science and policy. The authors provide some useful guidance by suggesting that we need to be more qualitative in how we model, for instance by recognizing that all current data and analysis point to sea levels rising for the foreseeable future, with some extreme scenarios that cannot be ruled out or even quantified probabilistically, but with a more honest admission that accurate prediction of long-term rates and totals is beyond our abilities.

Qualitative modelling supports adaptive decision-making under uncertainty, where commitments are tentative and continuously re-evaluated in the light of experience. Adaptive management does not exclude long-term planning, but it could help us to avoid the big mistakes made when we act as if we know more than we actually do. And as *Useless Arithmetic* details, making big mistakes based on the misuse of quantitative models is far more common than it should be. ■

Roger Pielke Jr is at the Center for Science and Technology Policy Research, University of Colorado/CIRES, Boulder, Colorado 80309, USA.

British Museum of Natural History in London over the ultimate repository of type specimens, which, thankfully, was resolved amicably.

Long's writing style is breezy and engaging, showing the skill of an experienced writer. If you ever wanted to know what can befall you on a fossil-collecting expedition, it's all here — everything from broken-down vehicles, to dingoes trailing you in hopeful expectation of your death, to sharing the remote and isolated collecting area with an armed serial killer on the run. Helpful marginal notes explain scientific concepts such as phylogeny (the science of family trees) and fieldwork trivia such as death adders (beautifully camouflaged and extremely venomous snakes).

The second section of the book continues in the same accessible vein with an overview of the scientific importance of the Gogo fossils and their place in the emerging story of early vertebrate evolution. This is not a comprehensive account — for that I would recommend Philippe Janvier's *Early Vertebrates* (Oxford University Press, 1996) — and different researchers in the field will no doubt have their own disagreements with parts of it, but for those new to the subject it is an excellent introduction. It also demonstrates how progress in vertebrate palaeontology depends on the discovery of top-quality specimens, fossils so well preserved that anatomical and functional questions can be answered without ambiguity.

Of course, there are quibbles: giant dragonflies are not known from the Devonian, and Erik Jarvik and Erik Stensiö were not directors of the Swedish Museum of Natural History, although they did head the palaeozoology section. There is also one serious omission:

Science set in stone

Swimming in Stone: The Amazing Gogo Fossils of the Kimberley

by John Long

Fremantle Arts Centre Press: 2007. 320 pp. \$26.95

Per E. Ahlberg

Palaeontology is essentially a science of serendipity. Most of biology rests on the planning and execution of experiments that may or may not work; palaeontologists plan and carry out expeditions to look for fossils that may or may not be there. Of course, this base of collected specimens supports an analytical superstructure as organized and strategically constructed as any other. But behind the meticulous published analyses there lurks the faintly thrilling — or unsettling — realization that a loose rock left unturned, or a lunch break taken in a different spot on the hillside, could have caused a specimen or even a whole locality of critical scientific importance to remain undiscovered.

In *Swimming in Stone*, John Long tells the story of one of these critically important localities: Gogo in Western Australia. Here, at Go Go Station near Fitzroy Crossing in the tropical north of the state, yellow limestone nodules weathering out on a plain backed by oddly abrupt hills yield 380-million-year-old fossil fishes of such extraordinary perfection that, in some specimens, the jaws can be smoothly opened and closed without grating on their hinges. The whole landscape is, in fact, a gigantic fossil: the hills were once a barrier reef with outlying atolls, and the black soil of the plain is the mud that accumulated on the fore-reef slope. Limestone nodules formed around animal carcasses in this mud, preserving them perfectly to the present day. The locality was discovered in 1940 when the first fossil fish was collected by geologist Curt Teichert, a refugee from Nazi Germany. Long, currently head of science at Museum Victoria in Melbourne, has been collecting fossils there for more than 20 years.

After opening with a brief scene-setting introduction, including an imagined dive on the Gogo reef, the first main section of the book charts the history of the site from its discovery, through collecting trips by British and Australian teams, to the period since the 1980s when Long became the driving force behind the collecting programme. Intertwined with this narrative of hard slog and lucky chance are vivid thumbnail portraits of the major players, such as Teichert, Harry Toombs and George Kendrick. There are also accounts of technical advances (crucially the development of fossil preparation using acetic acid, which revealed the full glory of the Gogo fishes) and scientific controversies over the Gogo material. These included not only issues of anatomical interpretation, but also a rather tense dispute between the Western Australian Museum in Perth and the then



nowhere can I find a discussion of why the Gogo fossils are so exceptionally well preserved. Nodular preservation of fishes is common in the Devonian, but the specimens are usually nowhere near as perfect as these. Any future second edition of the book should certainly be expanded to include a chapter on this topic. But in the end these are the minor blemishes of a delightful book.

My first ever contribution to *Nature*, 18 years

ago, was a News & Views piece on John Long's Gogo fish research (*Nature* 337, 511–512, 1989). I ended it by saying that the scientific value of the Gogo fossils is incalculable. This remains just as true today, and *Swimming in Stone* is a worthy testament to the fact. ■

Per E. Ahlberg is in the Department of Physiology and Developmental Biology, Uppsala University, Norbyvägen 18A, 752 36 Uppsala, Sweden.

piece of unexplained jargon. Sometimes we just need a convenient word to keep track of the main players." And with that, we stiffen, expecting the worst, but he goes on as before, escorting us through the deep forestal connections between mathematics and physics, guiding us with his inimitable clarity. But apart from some terms clear only to professionals, there's nothing to bear — we understand it.

When we come to Einstein, Stewart says: "Acres of print... have been devoted to Einstein's life and work." Yet when he tells us about Einstein's life and work, he makes us feel as if we've never heard it before. We may not fully understand the geometry of Minkowski space-time, the non-euclidean metrics or the bizarre consequences of relativity, but we come to believe we partly understand them — the ultimate aim of popular science narratives.

Why Beauty is Truth is a brilliant interweaving of politics, history and intrigue, with characters living ordinary lives, described in the spirit of a Russian novel.

An ode to symmetry

Why Beauty Is Truth: The History of Symmetry

by Ian Stewart

Basic Books: 2007. 304 pp. \$26.95

Joseph Mazur

Had Keats lived long enough to read Ian Stewart's latest book, *Why Beauty is Truth*, he might have been inspired to write a sixth great ode, soft-piping sweet melodies in praise of symmetry. But what did Keats mean in *Ode on a Grecian Urn* when he wrote some of the most famous lines in English poetry: "Beauty is truth, truth beauty — that is all/ Ye know on earth, and all ye need to know"?

Stewart is onto something deep, something mathematicians must have been keenly aware of since Greeks began turning urns. What is the underlying beauty of mathematics? Is it the artful way a proof is expressed? Or is it something deeper — something guiding pythagorean and platonist mathematicians to see better, something at the molecular structure of mathematics, some "unravish'd bride of quietness", some "Attic shape" — that enlightens and delights us.

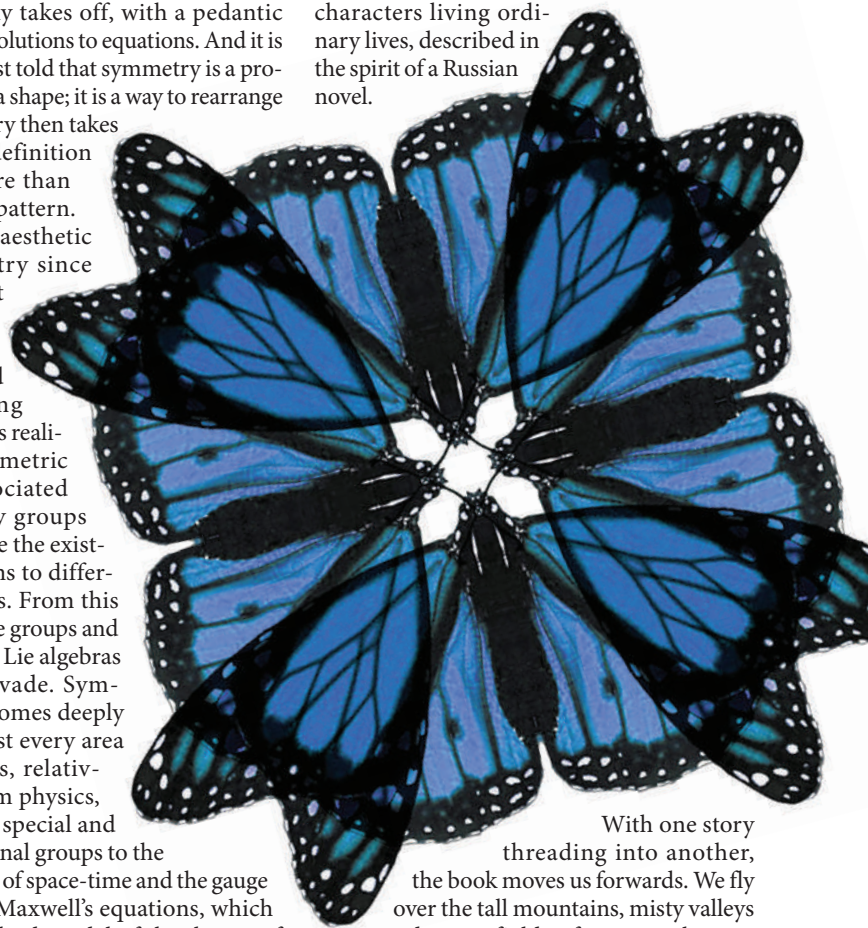
Stewart, a professor of mathematics at the University of Warwick, is renowned for his popular science books, but *Why Beauty is Truth* is without a doubt his finest. If it were just an authentic history of mathematics, it would be creditable. If it were only for its lively informal style, its historical characters, its intrigue ("The Galois group has a terrible secret"), its beautiful prose, it would be praiseworthy. Yet, its real uniqueness — its power — is in what it uncovers. It brings us the heart of why mathematicians pursue mathematics.

Beauty is not always as visible as the iridescent butterfly on the cover of Stewart's book. We are aware that it is not the dazzling colour that makes such an insect beautiful, but rather its shape, in particular its symmetry. It is this kind of beauty that Stewart's book reveals. We encounter it most obviously when we perceive it in geometry, in the wings of a butterfly, the sections of a cone, or the appearance of regular solids. But Stewart wants us to 'see' the invisible symmetries of algebra.

He starts with Evariste Galois, a young nineteenth-century French revolutionary who saw

them when he discovered — ostensibly on the night before he died in a duel — how symmetry could be used to decide which polynomial equations could be solved by radicals and which could not. This is where the story of symmetry really takes off, with a pedantic question about solutions to equations. And it is where we are first told that symmetry is a process rather than a shape; it is a way to rearrange things. Symmetry then takes on a rigorous definition to become more than just a pleasing pattern. It had been an aesthetic part of geometry since Pythagoras, but Stewart tells us that Galois' revelation inspired others, starting with Sophus Lie's realization that geometric invariants associated with symmetry groups could determine the existence of solutions to differential equations. From this step forward, Lie groups and their associated Lie algebras emerge and invade. Symmetry then becomes deeply rooted in almost every area of mathematics, relativity and quantum physics, from the use of special and unitary orthogonal groups to the transformations of space-time and the gauge symmetries of Maxwell's equations, which led to the standard model of the theory of all fundamental particles.

How does Stewart do it? How can he make such phenomenally difficult material so easy to understand without giving sophisticated readers a sense that he is oversimplifying, and at the same time create something that is such a joy to read? He uses tangible things as examples, describing the fundamental forces of nature by throwing dishes on a kitchen floor. Towards the end he gives us a warning, as if apologizing for pedantry. "Bear with me," he says, "and don't worry too much about the odd



With one story threading into another, the book moves us forwards. We fly over the tall mountains, misty valleys and green fields of current abstract maths and fundamental physics to witness the true beauties of truth. And in the end Stewart confesses: "No one could have predicted that a pedantic question about equations could reveal the deep structure of the physical world, but that is exactly what's happened."

Stewart has written an ode to Keats' equation so that we, too, can now witness the beauty of truth. ■

Joseph Mazur is the author of *The Motion Paradox*. He teaches mathematics at Marlboro College, Marlboro, Vermont 05344, USA.



The best is yet to come

Optimality is a key organizing principle of science, but the patterns of connections within real-world networks do not always respect it.

Mark Buchanan

An 'optimal' solution to a problem is, in some sense, the 'best' solution. It's the shortest route to work, or the way of packing oranges that takes up the least volume. In science, optimality has long been an organizing principle. Mathematical physics views the Universe as unfolding with dynamics that minimize a quantity known as the 'action', whereas economists and other social scientists often take optimality as a guide to human behaviour: we act, they say, to maximize our utility, be it financial or otherwise.

Looking at our networked and interconnected world, we may wonder whether anything is optimal here, from the food webs that underlie ecosystems, to technological networks such as the Internet. And science recognizes that the real world often falls short of optimality. Physicists know of many materials in which, as in glass, complex interactions among the molecules prevent the ordered arrangement of lowest (free) energy; these materials persist naturally in disordered, sub-optimal confusion. Meanwhile, the economists' *Homo economicus* has been replaced by a biologically more plausible creature acting on the basis of fast yet fallible instincts. We solve life's problems the way we pack the dishwasher — not optimally, after long calculation, but quickly and more or less efficiently.

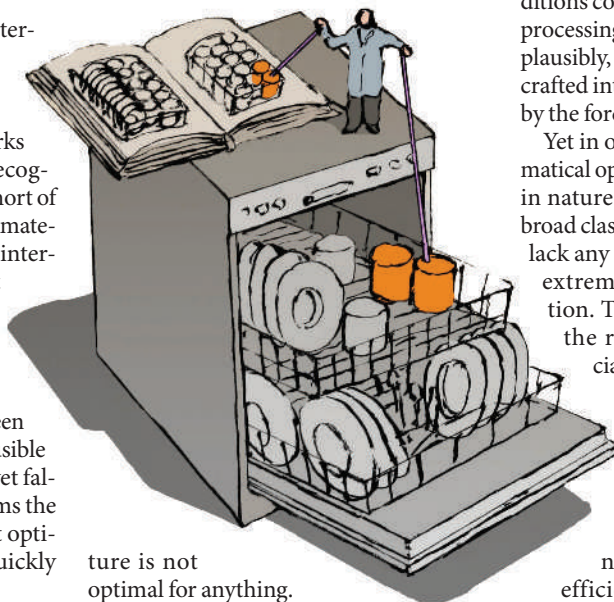
What about networks? Is the Internet optimal, in any sense? Is there a 'best' way to design the connections that link together a collection of cells, people or computers? This question has many possible answers — which always depend on what, if anything, a network is 'designed' to do.

In 1964, American engineer Paul Baran conceptualized a layout for the command and control network of the US military aiming to withstand an attack by the Soviet Union. He visualized it as a meshwork, something like a fishnet, with roughly the same number of links emanating from each element. No one element would be a communications hub — and primary target — and the natural redundancy of paths would allow messages to find a route through the network even if much of it was destroyed.

A few years later, US researchers had Baran's meshwork in mind while developing the ARPANET, the seed of today's Internet. Yet the Internet itself has turned out to be wildly different from Baran's

hubless ideal. Although decentralized like Baran's meshwork, it is in fact conspicuous for its communications hubs — network routers (or clusters of routers) that accumulate far more connections than most others. The network is especially sensitive to any failure of these centres. In terms of resilience, it is clearly not optimal.

Of course, the forces driving Internet growth have never been channelled towards resilience, and its architecture reflects the independent actions and decisions of untold millions of people and organizations. Quite possibly, its architect-



ture is not optimal for anything.

Yet in other settings the notion of network optimality may have considerable value. Suppose the population in a certain country has a known geographical distribution. How should facilities such as hospitals be located most conveniently? One recent study suggests a relatively simple, if somewhat peculiar, answer: that the density of facilities shouldn't just be proportional to that of people, but to that density raised to the power of two-thirds. This puts more facilities where there are more people, but not so many more that they become redundant. It is unlikely that urban planners currently adhere to this principle, but it presents a clear target for future policies.

Another problem that a network might solve is coordinating the activity of a set of elements — by helping them to synchronize their behaviour, for example. The elements could be living cells such as neurons, or anything else with rhythmic activity, and interact through signals of some

kind — electrical or chemical signals for cells, or light for fireflies in a tree. It turns out that the way such a network synchronizes partly reflects its community structure, a community being any cluster with far more links between its own elements than to elements elsewhere in the network. Studies show that under broad conditions, the smallest and densest communities tend to synchronize first, with larger-scale synchronization arriving later.

This suggests that a modular structure of communities within communities may allow for rich and varied dynamics — conditions conducive to efficient information processing and storage. In this sense, quite plausibly, the wiring of the brain has been crafted into optimal or near-optimal form by the forces of evolution.

Yet in other cases, even proven mathematical optimality may lack any exemplars in nature. Researchers have identified a broad class of networks that conspicuously lack any modular structure, yet support extremely efficient global coordination. These networks — named after the renowned Indian mathematician Srinivasa Ramanujan — have highly uniform architectures, with additional long loops lending them an 'entangled' character. They seem to be almost optimal for many purposes, for example by providing networks that can be easily and efficiently searched while avoiding congestion. Even so, no one as yet has found any natural network that adopts this style.

What all this suggests is that questions of optimality are rarely straightforward. A network will probably reveal optimal form only if it has a ready means for evolving towards it, and time to do so. And the nature of such optimality always depends on the network's function. In many cases today, even that remains at least a partial mystery.

Mark Buchanan is a science writer based in the United Kingdom.

FURTHER READING

Gastner, M. T. & Newman, M. E. J. *Phys. Rev.* **E74**, 016117 (2006).
Arenas, A., Diaz-Guilera, A. & Pérez-Vicente, P. J. *Phys. Rev. Lett.* **96**, 114102 (2006).
Donetti, L., Neri, F. & Muñoz, M. A. *J. Stat. Mech.* doi:10.1088/1742-5468/2006/08/P08007 (2006).

For other essays in this series, see <http://nature.com/nature/focus/arts/connections/index.html>

J. KAPUSTA

CONNECTIONS

NEWS & VIEWS

MATHEMATICS

A beauty and a beast

Hermann Nicolai

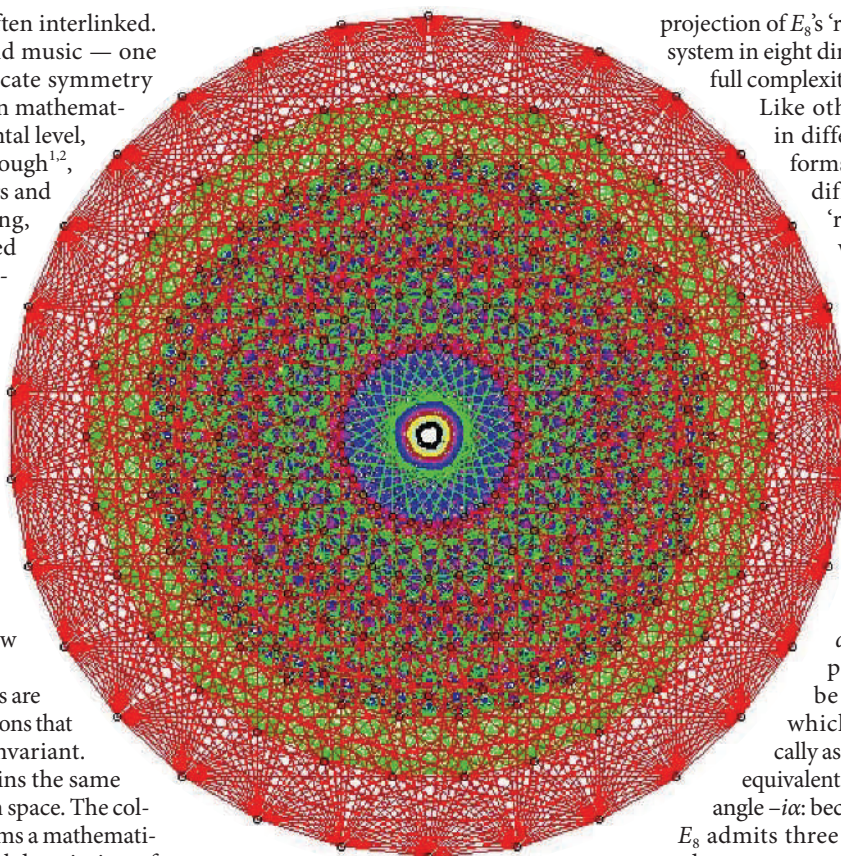
The mapping of the largest exceptional Lie group, E_8 , is a milestone for enthusiasts for the aesthetics of mathematics. But this embodiment of complex symmetry could be of interest to fundamental physics, too.

Symmetry and beauty are often interlinked. This is so not only in art and music — one need only think of the intricate symmetry of a Bach fugue — but also in mathematics and, at its most fundamental level, physics. In a recent breakthrough^{1,2}, which involved 18 researchers and was four years in the making, mathematicians have mapped out one of the most mysterious and fascinating of all mathematical objects: the ‘exceptional Lie group’ E_8 . In view of the magnitude of the computation and the sheer amount of data involved, the achievement has been likened to the mapping of the human genome³. But seeing the beauty in this complex beast can be hard: certainly more difficult than appreciating a Bach fugue without knowing the rules of counterpoint.

In mathematics, symmetries are usually associated with operations that leave a geometrical object invariant. A sphere, for instance, remains the same under continuous rotations in space. The collection of such operations forms a mathematical ‘group’. The mathematical description of continuous symmetries (as opposed to discrete symmetries, such as those that leave a crystal lattice invariant) is codified in the notion of a Lie group, named after the Norwegian mathematician Sophus Lie. Finite-dimensional Lie groups were classified more than a century ago, by Wilhelm Killing and Elie Cartan, by dint of considering only group elements infinitesimally close to identity: that is, to ‘rotations’ by arbitrarily small angles. Simply put³, they identified four infinite series of such groups, labelled A_n , B_n , C_n and D_n for $n = 1, 2, 3, \dots$, which essentially correspond to linear transformations in spaces of arbitrary dimension that leave certain quadratic expressions invariant. There are also five exceptional groups that do not fit into these categories, designated G_2 , F_4 , E_6 , E_7 and E_8 .

Visualizing rotations in three-dimensional

space is straightforward (as it is, with some training in mathematics, in higher dimensions!), but the ‘visualization’ of exceptional symmetries and their action on geometrical objects is much harder. The results of such attempts are often collectively (and jokingly) referred to as the ‘botany’ of these Lie groups. For instance G_2 , by far the ‘easiest’ of the exceptional groups, can be defined as the group that leaves invariant the multiplication table of a system of hypercomplex numbers known as octonions. E_8 stands out as the largest and most difficult of the exceptional Lie groups. It has 248 dimensions, and its smallest non-trivial realization requires a space of 57 dimensions^{1,2} (see ref. 4 for a physicist’s description of this object). In short, E_8 is as intricate as symmetry can get. Pictured here is a two-dimensional



projection of E_8 ’s ‘root system’ — a lattice-like system in eight dimensions that embodies its full complexity.

Like other Lie groups, E_8 comes in different versions, called real forms. Roughly speaking, these differ according to whether ‘rotations’ are performed with a real-number angle or an imaginary-number angle. More specifically, if it is possible to return to the starting point after a finite rotation, one speaks of a compact realization. A simple example is rotation in space by 360° , which can be represented mathematically through multiplication by $e^{i\alpha}$ with the (real) angle $\alpha = 2\pi$. A simple non-compact transformation would be translation along a line, which is realized mathematically as multiplication by e^α . This is equivalent to rotation by an imaginary angle $-i\alpha$: because $i^2 = -1$, then $e^{i(-i\alpha)} = e^\alpha$.

E_8 admits three real forms, one compact and two non-compact. Quite generally, the non-compact forms are much more tricky to deal with. This makes the main advance just reported¹ so impressive: it concerns the most subtle of all non-compact forms in Lie-group theory, the ‘split-real form’ of E_8 , sometimes denoted $E_{8(8)}$.

Aside from pure mathematics, what is the wider significance of this achievement? One answer lies in fundamental physics. Symmetry concepts played a central role in the establishment of the two most successful theories of modern physics: general relativity, and quantum-field theory as embodied in particle physics’ standard model. In general relativity, symmetry enters through the principle of general covariance: that the laws of physics should not depend on the coordinate system in which they are formulated. This principle enabled Albert Einstein to formulate in one

stroke the equations of the gravitational field governing the evolution of the Universe, as well as many other phenomena that would otherwise be intractable (the interaction of light with gravity, for instance).

In the standard model, symmetry is embodied by the principle of gauge invariance, which determines the way in which elementary particles can interact. Given this principle, and the apparatus of modern quantum-field theory, all that is needed to properly formulate the standard model is the specification of the symmetry group, the matter-particle content, and the transformation properties of these matter fields (quarks and leptons) under the chosen symmetry group. Gauge invariance automatically ensures the mathematical consistency ('renormalizability') of the theory, allowing us to extract definite predictions from seemingly infinite expressions, and thus making the standard model one of the best-tested theories of physics.

Yet in spite of their success, neither general relativity nor the standard model can be final theories of physics⁵. This is first of all because of a basic incompatibility between the two theories, reflected in the appearance of 'non-renormalizable' infinities when Einstein's theory is quantized following the standard rules of quantum mechanics. Equally importantly, neither theory is able to answer some obvious questions. For instance, what sets the pattern of elementary particles found in nature apart from other possible such patterns? Similarly, what is so special about the standard model's symmetry group, denoted $SU(3) \times SU(2) \times U(1)$, which seems mathematically undistinguished? And, connected to those questions, how did the Universe, and with it space-time and matter, come into being at the moment of the Big Bang?

To avoid the existing mathematical discrepancies, the yet-to-be-constructed unified theory (sometimes dubbed 'M theory') must be tightly constrained, and possibly even uniquely determined, by symmetry principles. One important difference between Einstein's theory of gravity and the standard model concerns the way in which symmetries are realized. In general relativity, symmetries act in physical space and time, whereas the gauge transformations of particle physics act in an abstract internal space (in which one can, for example, 'rotate' a proton into a neutron and vice versa).

An important step on the long road to a unified theory was the development of supersymmetry, a new kind of symmetry relating the particle groups known as bosons and fermions⁶. This led to supergravity, an extension of Einstein's theory, and superstring theory⁷, which is considered by many to be the leading contender to unify physics. Surprisingly, it turned out that the 'most supersymmetric' extension of Einstein's theory — supergravity in 11 space-time dimensions^{8,9} — has the split-real forms of E_6 , E_7 and E_8 automatically built into it¹⁰, albeit in a rather hidden form.

This seminal discovery was all the more

remarkable because it revealed completely unsuspected connections. Who could have anticipated what is, in effect, a link between the esoterics of exceptional Lie groups and the absence of long-range (tensor) forces other than gravity in nature? More recent studies of gauged maximal supergravity theories in three dimensions¹¹ have confirmed the intimate links between supergravity and the split-real form $E_{8(8)}$.

As yet, we have no idea what the true extent of E_8 's involvement in the scheme of things will be. If proponents of superstring theory are right, the compact form of E_8 could be realized as a gauge symmetry in the framework of 'grand unification'. But it is equally possible that E_8 will be realized in a different and more subtle way, intertwining space-time and matter, and possibly involving the split-real form, rather than the compact form.

The ambitious search for a fundamental symmetry of nature might even force us to venture into the unknown territory of infinite-dimensional exceptional symmetry groups, of which the finite-dimensional E_8 is just a subset. The

prime candidate is E_{10} , about which we know next to nothing, other than that it exists. Physicists should not let themselves get carried away by these intriguing possibilities, as experiment remains the final arbiter. But they would be well advised to take note of the exciting developments¹ in deciphering the E_8 group.

Hermann Nicolai is at the Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Mühlenberg 1, D-14476 Potsdam, Germany.
e-mail: nicolai@aei.mpg.de

1. www.liegroups.org
2. <http://aimath.org/E8>
3. Bourbaki, N. *Éléments de Mathématiques: Groupes et Algèbres de Lie: Chapitre 9* (Masson, Paris, 1982).
4. Günaydin, M., Koepsell, K. & Nicolai, H. *Comm. Math. Phys.* **221**, 57–76 (2001).
5. Ramond, P. *Journeys Beyond the Standard Model* (Perseus, New York, 1999).
6. Bagger, J. & Wess, J. *Supersymmetry and Supergravity* (Princeton Univ. Press, 1992).
7. Green, M. B., Schwarz, J. H. & Witten, E. *Superstring Theory* (Cambridge Univ. Press, 1987).
8. Nahm, W. *Nucl. Phys. B* **135**, 149–166 (1978).
9. Cremmer, E., Julia, B. & Scherk, J. *Phys. Lett. B* **76**, 409–412 (1978).
10. Cremmer, E. & Julia, B. *Nucl. Phys. B* **159**, 141–212 (1979).
11. Nicolai, H. & Samtleben, H. *J. High Energy Phys.* 04:022 (2001).

CHEMICAL BIOLOGY

Ignore the nonsense

Anton Schmitz and Michael Famulok

A small molecule forces the protein-translation machinery to overlook the signals that would otherwise result in its premature termination. Genuine stop signs are, however, read and obeyed.

Several inherited diseases are caused by mutations in single nucleotides within genes. These mutations can transform the products of messenger RNA codons, the sets of three nucleotides that determine which amino acid is incorporated into the growing protein chain. When such 'nonsense' mutations are transcribed into a 'stop' codon, the cellular machinery that translates mRNA into protein misinterprets the codon as a signal to terminate protein synthesis. These false stop codons are known as premature termination codons (PTCs) and result in the formation of truncated proteins that cannot function properly and may even damage the cell, eventually leading to disease. Depending on the disorder, nonsense mutations account for 5–70% of cases of genetic disorders, including cystic fibrosis, muscular dystrophy and several types of cancer. On page 87 of this issue, Welch *et al.*¹ report that a small organic molecule known as PTC124 can force the translation machinery to ignore PTCs, without preventing it from reading the real stop signals*.

It has been known for the past 10 years that the antibiotic gentamycin can prompt

ribosomes — the core component of the cellular protein-synthesis machinery — to read through PTCs, thereby generating full-length proteins². Nevertheless, the clinical benefit of gentamycin is limited, because to be effective it has to be used at very high concentrations, which are associated with severe side effects. There is now hope that PTC124, which, like gentamycin, ignores PTCs but lacks its adverse side effects, could be more beneficial in the clinic. Indeed, interim results of phase II clinical trials³ indicate that patients with PTC-induced forms of cystic fibrosis and Duchenne muscular dystrophy might benefit from treatment with PTC124 — a promising result that has been commented on for some time^{4,5}.

Welch *et al.*¹ describe an astonishing feature of PTC124 — its selectivity for PTCs. Why is this so striking? All organisms with membrane-bound cell nuclei (eukaryotes) have evolved mechanisms to protect themselves from the harmful products of nonsense mutations. There are two lines of defence. The first relies on the fast and efficient degradation of the truncated proteins after the translation of PTC-containing mRNAs. The second acts before these proteins are synthesized. This quality-control mechanism, known as

*This article and the paper concerned¹ were published online on 22 April 2007.

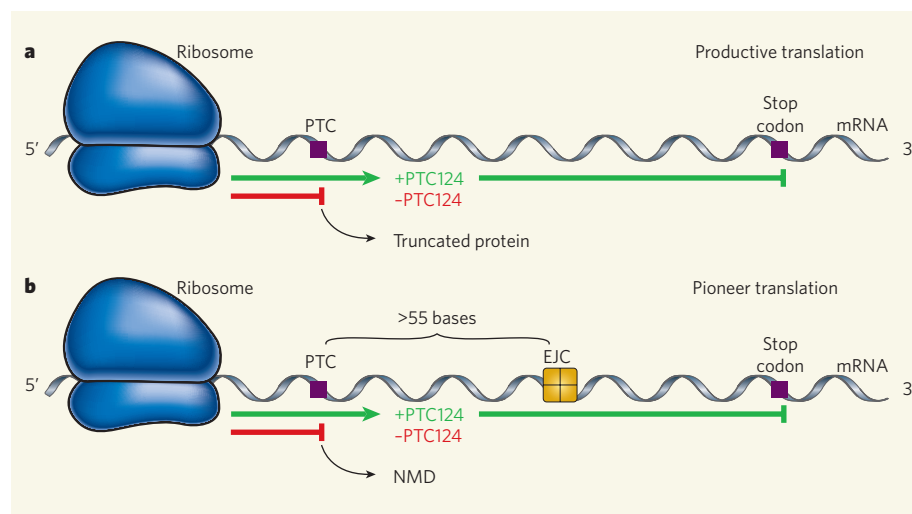


Figure 1 | Possible mechanisms of PTC124-enhanced translation. **a**, PTC124 could directly suppress termination of 'productive' protein translation at a premature termination codon (PTC), leading to increased levels of functional full-length proteins. **b**, Alternatively, PTC124 might suppress the identification of PTCs in 'pioneer' translation, thereby preventing nonsense-mediated mRNA decay (NMD), which is induced if a ribosome hits a stop codon more than 55 bases upstream of an exon-junction complex (EJC). This would lead to the stabilization of PTC-containing mRNAs, allowing their translation, presumably through the mechanism shown in **a**. The work of Welch *et al.*¹ supports the model depicted in **a**.

nonsense-mediated mRNA decay (NMD), detects and degrades PTC-containing mRNAs⁶. But how is a PTC distinguished from a normal termination codon? The answer came from the discovery that, in all organisms, NMD requires ongoing protein translation.

The only cellular entity that can read the genetic code, and so decipher stop codons, is the ribosome. In mammals, PTC recognition occurs during an initial 'pioneer' round of translation in which — unlike 'productive' translation, which results in protein synthesis — a ribosome scans only newly synthesized mRNAs for the presence of PTCs⁷. NMD is closely linked to mRNA splicing — a process in which non-coding sequences of a gene (introns), which have been transcribed into the precursor mRNA, are removed. After re-joining of the coding sequences (exons), the splicing machinery labels the joining sites by attaching a set of proteins known collectively as the exon-junction complex. As a rule, the ribosome identifies a termination codon as premature if it is more than 55 bases upstream of an exon-junction complex⁸.

Although NMD's role in mRNA quality control has been known for many years, recent observations indicate that this might not be its main function. On inactivation of the NMD machinery, about 10% of the transcriptome — the set of all mRNAs in a cell — of yeast and human cells are differentially expressed, indicating that NMD is involved in the regulation of gene expression^{9–11}. Examples of NMD-regulated molecules include several proteins that are involved in amino-acid synthesis, as well as pseudo-genes that do not encode a protein but are the source of non-coding RNAs such as small nucleolar RNAs¹².

Therefore, drug-induced readthrough of

PTCs could enhance protein expression in two ways. First, by directly suppressing premature termination during the productive translation of mRNA, it could increase the efficiency of translation (Fig. 1a). Second, by suppressing PTC recognition during pioneer translation, it could prevent NMD, thereby increasing the number of mRNA sequences that are available for productive translation (Fig. 1b). Although both possibilities lead to a favourable increase in protein translation, the second is associated with a risk of an increased amount of mutated proteins and mRNAs, which are normally absent because of mRNA degradation by NMD.

Welch *et al.*¹ found that, fortunately, PTC124

seems to selectively target productively translating ribosomes, and does not interfere with those that are engaged in NMD. In this way, PTC124 differs from gentamycin, which induces upregulation of several mRNAs that are known to be NMD substrates.

As well as being a prerequisite for its therapeutic efficacy, the specificity of PTC124 could be a starting point for a better understanding of the differences between the productive and the pioneer modes of translation. Although other distinct agents (besides ribosomes) are required for these two types of translation, we don't have detailed structural and functional information about them. The identification of the molecular target of PTC124 should help to fill this gap. Thus, in addition to being a promising drug, PTC124 may also become a research tool for elucidating the underlying mechanisms of the termination of protein translation.

Anton Schmitz and Michael Famulok are at the LIMES Institute, Program Unit Chemical Biology and Medicinal Chemistry, University of Bonn, Gerhard-Domagk-Strasse 1, 53121 Bonn, Germany.
e-mail: m.famulok@uni-bonn.de

1. Welch, E. M. *et al. Nature* **447**, 87–91 (2007).
2. Howard, M., Frizzell, R. A. & Bedwell, D. M. *Nature Med.* **2**, 467–469 (1996).
3. www.ptc124.com/3.11_generic_disorders.aspx
4. Ainsworth, C. *Nature* **438**, 726–728 (2005).
5. Hamed, S. A. *Drugs* **9**, 783–789 (2006).
6. Chang, Y.-F., Imam, J. S. & Wilkinson, M. F. *Annu. Rev. Biochem.* **76**, doi:10.1146/annurev.biochem.76.050106.093909 (2007).
7. Ishigaki, Y., Li, X., Serin, G. & Maquat, L. E. *Cell* **106**, 607–617 (2001).
8. Nagy, E. & Maquat, L. E. *Trends Biochem. Sci.* **23**, 198–199 (1998).
9. Lelivelt, M. J. & Culbertson, M. R. *Mol. Cell. Biol.* **19**, 6710–6719 (1999).
10. He, F. *et al. Mol. Cell* **12**, 1439–1452 (2003).
11. Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. *Nature Genet.* **36**, 1073–1078 (2004).
12. Mitrovich, Q. M. & Anderson, P. *Curr. Biol.* **15**, 963–967 (2005).

PARTICLE PHYSICS

Wobbly oscillations

David Wark

Neutrinos seem to oscillate: they change back and forth between one type and another and, by extension, have a tiny mass. But one experiment that predicted a particularly large mass looks to have been mistaken.

When is a discovery not a discovery? When it can't be reproduced, of course. That scientific ground-rule has plagued the members of the LSND (Liquid Scintillator Neutrino Detector) collaboration since they first saw evidence¹ for so-called neutrino oscillations. Had the LSND results been confirmed, they would have rewritten much of what we think we know about the ever-elusive neutrinos (Box 1, overleaf). But results just announced² from the MiniBooNE detector at Fermilab, near Chicago,

could prove those earlier results' nemesis.

The story begins in 1996, with observations made by LSND of the decay products of a pion particle beam at the Los Alamos Meson Physics Facility (LAMPF) accelerator in New Mexico. The neutrinos came mainly from the decay of positively charged pions into positive muons and muon neutrinos, and the subsequent decay of these positive muons to positrons (positively charged electrons), muon antineutrinos and electron neutrinos (Fig. 1a, overleaf). What

Box 1 | Neutrinos and their oscillations: a short history

Neutrinos belong to the class of elementary particles known as leptons. The electron is the most familiar lepton, but is merely a member of the lightest of — according to the standard model of particle physics — three families. Equivalent to the electron in every way except its mass (which is more than 200 times greater than the electron's 0.511 megaelectronvolts) is the muon; nearly 3,500 times as massive is the monstrous tau. Each of these three has its antimatter equivalent, which has an opposite electric charge of +1. The six charged leptons each have an uncharged, almost massless counterpart — the three neutrinos, and three antineutrinos.

Whether neutrinos do have mass and, if so, how much of it, is of considerable interest both to particle physicists, for the development of 'grand unified

theories', and to cosmologists. Neutrinos are one component of the invisible 'dark matter' that makes up roughly a third of the Universe's mass, and their mass has a significant effect on how the matter in the Universe is distributed.

If neutrinos do have mass, then it is possible that neutrino states will oscillate. An electron neutrino, for instance, might spontaneously change into a muon neutrino, and back again, in a regular rhythm. The larger the difference in neutrino masses, the shorter the distance over which this will happen. The idea was originally invented to explain why, in pioneering measurements made by Ray Davis in the 1960s, far fewer neutrinos were seen coming from the Sun than models of the nuclear-fusion processes taking place there demanded.

In 1998, results from the Super Kamiokande detector in Japan⁴ showed that the flux of neutrinos resulting from collisions of cosmic rays with Earth's atmosphere varies with direction. This convinced most physicists that neutrino oscillations had clearly been seen. In 2001, the Sudbury Neutrino Observatory in Ontario, Canada, showed⁵ that the Sun emits neutrinos other than electron neutrinos, which is very difficult to explain except through oscillations.

These results were further strengthened by the KamLAND experiment, which observed⁶ the apparent disappearance of electron antineutrinos from power reactors in Japan as they oscillated into muon and tau antineutrinos, which the detector could not see. The K2K (ref. 7) and MINOS (ref. 8) experiments saw

a similar apparent disappearance of muon neutrinos from beams that had propagated for hundreds of kilometres through the ground. Collectively, these experiments have built a case for the oscillation of atmospheric and solar neutrinos that looks almost bullet-proof. These oscillations determine the two independent mass differences that exist between the three neutrino types; both are found to be much less than 1 electronvolt.

The results from five years of data-taking at the LSND experiment¹ were distinguished by finding a much larger mass difference of around 1 electronvolt. If confirmed, this would wreck the simple explanation of the other experiments in terms of neutrino oscillations, and require much more complicated physics. But is it right? The MiniBooNE experiment aimed to find out. **D.W.**

the LSND collaboration found on examining the reaction products was an excess of electron antineutrinos¹ — which are produced nowhere in the positive-pion decay chain. The conclusion was that muon antineutrinos were changing into electron antineutrinos while propagating. This is the process known as neutrino oscillation, and is itself by now relatively uncontroversial³. But the oscillations that LSND saw seemed to indicate much larger neutrino mass differences than other experiments had predicted. That was indeed controversial.

Perhaps the LSND results pointed to some new physics even more surprising than neutrino oscillations. Many possibilities have been suggested, including the existence of other, 'sterile' neutrino families beyond the three that participate in the interactions of particle physics' standard model. But the first attempt to test the results — KARMEN, a comparable, but slightly less sensitive experiment at the Rutherford Appleton Laboratory in Didcot, England — failed to see any signal, although it could not rule out the LSND claim. A new experiment with much higher rates, and therefore higher sensitivity, was needed to test the LSND claim conclusively.

Thus was born Fermilab's MiniBooNE, so called because it was a scaled-down version of the proposed Booster Neutrino Experiment. In MiniBooNE, a beam of either positive or negative pions decays to form beams of positive or negative muons, and muon neutrinos or antineutrinos, whose oscillations are looked for (Fig. 1b). But the muons also decay, producing, among other things, electron neutrinos and antineutrinos — a potential source of confusion for the oscillation measurement. This is minimized at MiniBooNE by stopping the muons

before significant numbers of them decay, and also — unlike the LSND and KARMEN experiments — using a higher-energy pion beam, which produces a clean differentiation in energy between the different sources of neutrinos.

The muon neutrino beam propagates for around 500 metres before hitting a detector consisting of 1,520 light-sensitive photomultiplier tubes embedded in a 12-m-diameter tank of mineral oil. The charged particles resulting from neutrino interactions travel faster than the speed of light in the oil, resulting in the emission of a characteristic cone of 'Čerenkov' light, rather like the sonic boom produced by a

jet breaking the sound barrier. The detection of this light allows the particle's energy and position to be measured, and one type of neutrino to be distinguished from another through its characteristic interactions (Fig. 2).

The analysis of the data² is complicated, partly owing to physics and partly to psychology. The physics complication lies in the number of background signals that mimic the real one. Added together, these outnumber the postulated 'real' events by two to one. To add to the fun, neutrino cross-sections — the probability that a neutrino will interact in a particular way — are poorly known in the MiniBooNE energy range.

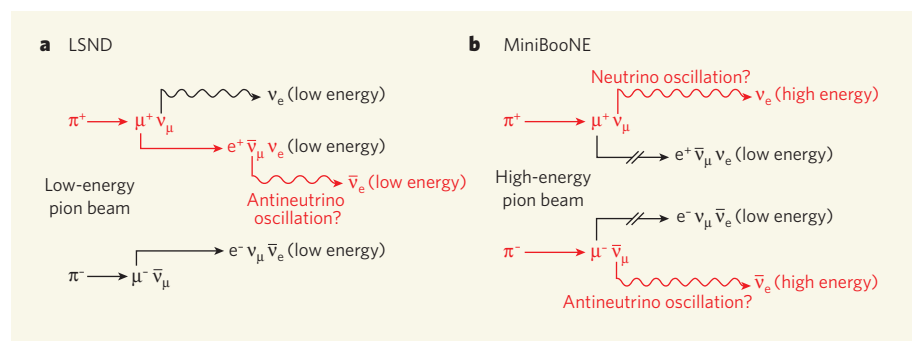


Figure 1 | Cleaning up the oscillation signal. **a**, The controversial LSND evidence for antineutrino oscillation used a low-energy beam of mainly positively charged pions (π^+). These decay into positive muons (μ^+) and muon neutrinos (ν_μ); the positive muons decay further into positrons (e^+), muon antineutrinos ($\bar{\nu}_\mu$) and electron neutrinos (ν_e). The presence of these last electron neutrinos made it difficult to obtain a clean measurement of the oscillation of ν_μ directly produced in the initial pion decay; instead, the oscillation of $\bar{\nu}_\mu$ from the secondary decay into electron antineutrinos ($\bar{\nu}_e$) was investigated. A further potential source of confusion is the presence of a small negative-pion (π^-) component in the initial pion beam. These pions decay just as do positive pions, but conjugately — positive particles are swapped for negative, and antineutrinos for neutrinos, and so on — thus also producing a small number of $\bar{\nu}_e$. **b**, The MiniBooNE experiment uses a high-energy pion beam, and the muons produced in the pion decays are mostly stopped before they can decay further. Oscillation of neutrinos directly produced in the π^+ decay can thus be measured cleanly, as these neutrinos are distinguished by their higher energy. Equally, oscillations of antineutrinos from π^- decay can be seen, albeit at a slower rate — this decay channel remains to be investigated.

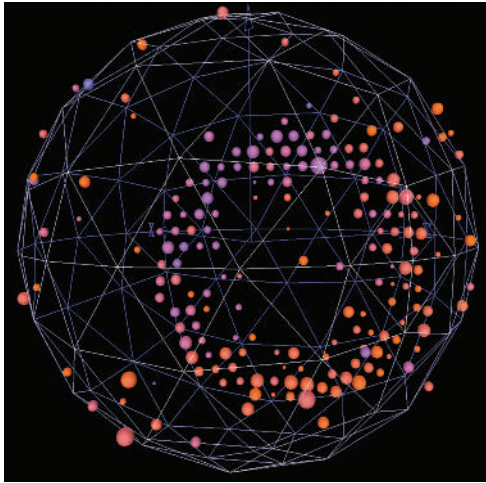


Figure 2 | Light signal. A neutrino interaction with the mineral oil in the 12-metre-diameter tank of the MiniBooNE experiment gives out light that fires photomultiplier detectors embedded in its sides in a pattern characteristic of the interaction products — and thus of the original type of neutrino. The circle of photomultiplier hits indicates a characteristic cone of ‘Čerenkov’ light emitted by a particle travelling at more than the speed of light in the medium (colours indicate the relative timing of the hits). With this apparatus, MiniBooNE was unable to confirm the evidence of neutrino oscillation found earlier in the LSND experiment.

Psychology enters the picture in the attempts to quantify the background signal. The history of science is littered with false results produced by scientists who see how each tuning of an analysis procedure affects their desired signal and its undesired background, and who subconsciously tune their analysis to produce the result they want to see. To prevent this, MiniBooNE adopted two independent forms of blind analysis, hiding away the data that would contain the potential signal until all the analyses had been finalized.

This procedure greatly increases the reliability of the results, but it also adds considerably to the complexity of the analysis. The two separate analyses reveal that the number of events corresponding to electron neutrinos matches

expectations from backgrounds. One analysis produces a few more events than expected, the other a few less, but neither deviation is statistically significant. The results thus rule out the simple oscillation explanation of the LSND results at the 98% confidence level.

The work is not yet done, however: 98% is a little less confidence than we like to have in our fundamental laws. MiniBooNE is working to increase its sensitivity still further, and also to understand a discrepancy between the expectation and experiment that occurs mostly at energies below those at which the oscillation signal should occur. In addition, a bolt-hole for the new physics that the LSND result might have signalled remains to be closed: LSND used antineutrinos; MiniBooNE, neutrinos. Such

a gross difference in the oscillations of these matter and antimatter analogues would be a truly revolutionary result. But no experimental results rule it out: experimenters must therefore try to collect enough data with an antineutrino beam to discount this possibility.

None of these quibbles seems strong enough to challenge the basic conclusion. It is worth mentioning here that the MiniBooNE collaboration includes many influential members of the LSND collaboration. They would not be human if they didn't have a strong desire to see their signal confirmed and most of neutrino physics rewritten: various rewards await those who rewrite the laws of physics. And yet they helped construct an experiment to produce a conclusive test of their results, intentionally designed it so that it would be almost impossible to bias the results one way or the other, and when it ruled against them, announced it openly to the world. That might not win a Nobel prize, but it is still science at its best. ■

David Wark is in the Department of Physics, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2BW, and at the STFC Rutherford Appleton Laboratory, UK. e-mail: d.l.wark@rl.ac.uk

1. Aguilar, A. *et al.* *Phys. Rev. D* **64**, 112007 (2001).
2. Aguilar-Arevalo, A. A. *et al.* preprint available at <http://arxiv.org/abs/0704.1500v2> (2007).
3. Giunti, C. *J. Phys. G* **34**, R93–R109 (2007).
4. Fukuda, Y. *et al.* *Phys. Rev. Lett.* **81**, 1562–1567 (1998).
5. Ahmad, Q. R. *et al.* *Phys. Rev. Lett.* **89**, 011301 (2002).
6. Eguchi, K. *et al.* *Phys. Rev. Lett.* **90**, 021802 (2003).
7. Ahn, M. H. *et al.* *Phys. Rev. D* **74**, 072003 (2006).
8. Michael, D. B. *et al.* *Phys. Rev. Lett.* **97**, 191801 (2006).

NEUROSCIENCE

Unconscious networking

Mark A. Pinsky and Sabine Kastner

What are neural networks doing when the brain is at rest? It turns out that in primates, even under conditions of deep anaesthesia, some of these networks undergo highly organized patterns of activity.

Our brains use up enormous amounts of energy, even when we are daydreaming with our eyes closed and not performing any demanding mental operations¹. In fact, intensive cognitive operations — arithmetic calculations, for example — increase the brain's energy consumption only minimally.

When subjects are in a resting state, spontaneous brain activity is not as chaotic as one might expect. Instead, that activity correlates systematically across anatomically and functionally connected areas that are normally used when performing tasks such as reading this article². The significance of this correlated activity during states of rest has remained unclear. Vincent and colleagues (page 83 of this issue)³ shed light on the matter by showing that organized activity patterns in neural networks

are similar across primate species, and are not tied to a conscious state of mind.

Vincent *et al.* used functional magnetic resonance imaging (fMRI) to investigate spontaneous fluctuations of neural activity in the monkey brain during anaesthesia. The fMRI signals are indirect measures of neural activity, and determine spatially specific changes in blood oxygenation levels across the brain (referred to as the BOLD signal). The authors chose a starting point, or ‘seed’ region, in the frontal cortex known as the frontal eye field, and looked at how spontaneous fluctuations of fMRI signals in this region correlated over time with signals in the rest of the brain.

The frontal eye field is part of the oculomotor system, a network of brain areas that subserves the planning and execution of eye

movements, and it is well understood in both monkeys and humans⁴. Only a few other discrete regions in the frontal and parietal cortex showed temporally coherent correlations with the spontaneous signal fluctuations in the frontal eye field. These brain regions are known to be interconnected, and are all part of the oculomotor system. When a different seed region in the oculomotor system was chosen, the same discrete network was revealed.

How do the activations across the oculomotor network, under conditions of deep anaesthesia, compare with those observed in awake monkeys performing eye movements? Strikingly, the authors found activations in the same brain regions in monkeys trained to perform eye movements in fMRI experiments⁵. Thus, the same network that is used during performance of the task maintains a state of correlated activity in the resting, and even unconscious, brain.

The oculomotor system is not the only brain network that shows organized patterns of signal fluctuations in anaesthetized monkeys. Vincent *et al.* observed correlated signal fluctuations in two other systems — the somatosensory/motor (somatomotor) system, which is involved in movement and touch, and the visual system.

When the authors used the somatomotor

cortex in one brain hemisphere as a seed region, they found that its BOLD signals correlated highly only with the somatomotor cortex in the opposite hemisphere. Interestingly, this region did not show high signal correlations with nearby frontal regions of the oculomotor system. This suggests that these correlations are network specific, and do not spread into neighbouring, but functionally distinct, regions of the cortex. In the visual system, a seed region in the primary visual cortex was chosen that represents a specific part of visual space. Remarkably, the fluctuations in BOLD signals correlated only with signals of subregions within other visual areas that represent the same part of visual space. Together, these findings show that the spontaneous signal correlations in the anaesthetized monkey are highly specific, both across functionally defined networks and within topographically defined subregions of a network.

Previous studies have shown that the main human cortical networks exhibit correlated spontaneous activity while subjects are at rest^{2,6,7}. Vincent and colleagues provide the first evidence that such activity is neither restricted to the human brain nor tied to a conscious state. Their findings suggest that fluctuations of spontaneous activity across anatomically interconnected brain regions constitute a fundamental principle of brain organization. Such an interpretation is supported by the fact that organized patterns of brain activity are present in both humans and non-human primates.

As to the functional significance of correlated signal fluctuations, it may be that they maintain the integrity of the networks by reinforcing the synaptic connections between neurons that are essential for network operations in the awake state. Indeed, in stroke patients, the functional connectivity of a brain network has been found to break down when one of its parts is damaged⁸. This loss of connectivity seemed to be correlated with the patients' behavioural impairments. Thus, the new findings³ may help in understanding both normal and pathological brain function.

Vincent *et al.*³ also investigated a possible monkey homologue of a cortical network that thus far has been studied only in humans. This human 'default' network exhibits BOLD activations when subjects are not performing any particular task, and is thought to support uniquely human functions — for example, thinking about ourselves and others, imagining the future, and daydreaming^{9–11}. The authors chose to study a seed region in the posterior cingulate cortex of the monkey brain; this brain region is anatomically similar in both species and is part of the human default network. They identified correlated activity in discrete regions of the frontal, parietal and temporal cortex, which may thus form an analogous default network in the monkey brain.

These findings challenge the view that the default network is uniquely human and is tied to human mental capabilities. But that

challenge depends on the assumption that the posterior cingulate cortex is analogous in both species: despite the anatomical similarities, it is not known whether this area serves similar brain functions in the two species. Furthermore, the human default network has been defined in the awake state, whereas this possible monkey homologue was investigated under deep anaesthesia. Further investigations of this network in the monkey brain, under conditions similar to those used in the human studies, will be necessary to clarify its relation to the human default network.

Vincent and colleagues' results³ raise other fascinating issues. How many brain networks can be discovered by identifying correlations between spontaneous signals? Also, what are the physiological criteria that define such networks? Arguably, the brain regions that form a network have to be interconnected. But as Vincent *et al.* show, network activations extend beyond those expected simply from connections between single synapses. Furthermore, are these organized patterns of activity unique to primates? The authors' approach may prove promising in revealing similarities between

brain regions in different species, especially when combined with task-induced brain activations that are related to sensory or cognitive functions. ■

Mark A. Pinsky and Sabine Kastner are in the Department of Psychology and the Center for the Study of Brain, Mind and Behavior, Green Hall, Princeton University, Princeton, New Jersey 08544-1010, USA.
e-mail: skastner@princeton.edu

1. Raichle, M. E. & Mintun, M. A. *Annu. Rev. Neurosci.* **29**, 449–476 (2006).
2. Biswal, B. B., Kylene, J. V. & Hyde, J. S. *NMR Biomed.* **10**, 165–170 (1997).
3. Vincent, J. L. *et al.* *Nature* **447**, 83–86 (2007).
4. Wurtz, R. H. & Goldberg, M. E. *The Neurobiology of Saccadic Eye Movements* (Elsevier, New York, 1989).
5. Baker, J. T., Patel, G. H., Corbetta, M. & Snyder, L. H. *Cereb. Cortex* **16**, 447–459 (2006).
6. Vincent, J. L. *et al.* *J. Neurophysiol.* **96**, 3517–3531 (2006).
7. Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L. & Raichle, M. E. *Proc. Natl Acad. Sci. USA* **103**, 10046–10051 (2006).
8. He, B. J. *et al.* *Neuron* **53**, 905–918 (2007).
9. Vogeley, K. & Fink, G. R. *Trends Cogn. Sci.* **7**, 38–42 (2003).
10. Gusnard, D. A., Akbudak, E., Shulman, G. L. & Raichle, M. E. *Proc. Natl Acad. Sci. USA* **98**, 4259–4264 (2001).
11. Mason, M. F. *et al.* *Science* **315**, 393–395 (2007).

NUCLEAR CHEMISTRY

Panning for ununbium

Andreas Türler

The chemical identification of two atoms of element 112 — scooped from the helium stream they were suspended in using a gold pan — brings the superheavy elements' fabled island of stability into sharper focus.

The superheavy element 112, provisionally named ununbium, is expected to belong to group 12 of the periodic table, which includes the familiar transition metals zinc, cadmium and mercury. As a metal, mercury is already quite unusual. It is volatile — it is liquid at room temperature — and chemically rather inert. A further peculiarity is its ability to form alloys, known as amalgams, with many metals. For that reason, miners traditionally used mercury to recover gold from its ores (Fig. 1, overleaf). On page 72 of this issue, Eichler *et al.*¹ turn the tables, using gold to pan for mercury's heavier homologue, element 112. Against all odds, they struck rich. Their haul was a mere two atoms that decayed again within a few seconds, but the consequences are nonetheless far-reaching.

Superheavy elements owe their existence solely to so-called nuclear-shell effects²: certain 'magic' combinations of many protons and neutrons arrange themselves so favourably that they can act against the disruptive forces of the protons' charges. Nuclei with these configurations are, unlike many heavy nuclei, fairly stable against spontaneous fission. The most stable nuclei might survive up to years — an

effect comparable to the chemical inertness of the noble gases elsewhere in the periodic table. Such nuclei are thought to inhabit an 'island of stability' conjectured to lie on the horizon of the current periodic table.

Shortly after physicists postulated the existence of superheavy elements in the 1960s, chemists joined in the rush to prospect for them. This not only involved a frantic search for these elusive elements in nature, but also led to the construction of new accelerators and the development of fast online detection techniques³. The euphoria was fuelled by rather exotic predictions: "Are elements 112, 114 and 118 relatively inert gases?" asked the renowned theoretical chemist Kenneth Pitzer in 1975 (ref. 4). Owing to the superheavy atoms' very large nuclear charge, some electrons in them move with velocities close to the speed of light, and must be treated relativistically. This relativistic correction might significantly alter the order of the elements' electronic orbitals, and so severely limit their chemical reactivity. The superheavy elements thus developed into a sort of sanctuary for theoretical chemists: any prediction made there was comparatively safe from experimental verification.



Figure 1 | Element panning, old-school style.

But this situation has been changing. The synthesis of several superheavy nuclei in fusion reactions involving heavy ions has been reported recently⁵. These spectacular results had, however, been questioned because other groups had failed to reproduce them. Eichler *et al.*¹ have just succeeded — at least in one case.

In an experiment lasting several weeks, the authors, working at the Flerov Laboratory of the Joint Institute for Nuclear Research in Dubna, Russia, bombarded a target of the plutonium isotope ²⁴²Pu with an intense beam of calcium (⁴⁸Ca) ions. Their hope was that two nuclei would occasionally fuse to form the superheavy isotope ²⁸⁷114, after the emission of three neutrons. From previous experiments at the same laboratory, ²⁸⁷114 was known to decay within about half a second by the emission of an α -particle (a helium nucleus, ⁴He, of two protons and two neutrons) to ²⁸³112.

Owing to the relatively long half-life (several seconds) of ²⁸³112, this nucleus can be trapped in helium gas and thus swept to a detector consisting of a narrow channel formed of gold-covered silicon detectors. To cover for the eventuality that element 112 is highly volatile (as, for example, the heavy noble gas radon is), the authors cooled the far end of this channel to -180°C . Under these same conditions, a diffusion-led process would deposit radioactive mercury isotopes on to the gold surface near the warm entrance of the detector; the still-more-volatile radon isotopes, by contrast, would be collected near the cold end.

So, would single atoms of element 112 behave in a manner becoming of their class — in other words, similarly to mercury, as predicted by new, very elaborate theoretical predictions⁶? Or would they act more exotically — that is, like radon, as speculated by Pitzer⁴? The former, say Eichler and colleagues, who observed two unique mercury-like signals near the warm end of their detector. First, they saw an α -particle with an energy of 9.5 mega-electronvolts, characteristic of the decay of

²⁸³112 to the darmstadtium isotope ²⁷⁹Ds; the signal was followed shortly after by the signal of this isotope's spontaneous fission.

The observations of long-lived α -decay of ²⁸³112 confirm that the island of stability is not just a mirage. Indeed, the ground on its neutron-poor shores is seeming ever more solid. As the synthesis of relatively long-lived isotopes of element 114 in reactions of ²⁴⁴Pu and ⁴⁸Ca has been reported, the next step should be to perform first experiments with this element. After all, element 112 seems to behave chemically quite 'normally'; that is, similarly to its lighter group partner mercury — although it could be that 112 is considerably more inert and volatile. The current experiment was not designed to investigate that, and further observations are needed.

So should we resume searching for superheavy elements in nature? Perhaps: armed with experiments such as those of Eichler and colleagues¹, we at least have a much better idea what to look for. Like the authors, we might, against all odds, strike it rich. ■

Andreas Türler is at the Institut für Radiochemie, Technische Universität München, Walther-Meissner-Straße 3, D-85748 Garching, Germany
e-mail: andreas.tuerler@radiochemie.de

1. Eichler, R. *et al. Nature* **447**, 72–75 (2007).
2. Stoyer, M. A. *Nature* **442**, 876–877 (2006).
3. Herrmann, G. in *The Chemistry of Superheavy Elements* (ed. Schädel, M.) 291–318 (Kluwer Academic, Dordrecht, 2003).
4. Pitzer, K. S. *J. Chem. Phys.* **63**, 1032–1033 (1975).
5. Oganessian, Yu. Ts. *J. Phys. G* **34**, R165–R242 (2007).
6. Sarpe-Tudoran, C. *et al. Eur. Phys. J. D* **24**, 65–67 (2003).

EARTHQUAKES

Relationships in a slow slip

Heidi Houston and John E. Vidale

The size and duration of disparate, slow, low-amplitude earthquake processes seem to obey a single scaling law. The relationship is very different from that which governs their more violent and impulsive cousins.

Subduction zones — those regions of Earth's crust where one tectonic plate dives beneath another — are usually associated with frequent and violent earthquake activity. But not always. Occasionally, the downgoing tectonic plate lurches slowly, smoothly and almost silently under the overriding plate, accompanied by weak bursts of tremors^{1,2}. These subtle episodes of tremors and slow slip, low in amplitude (and therefore only recently discovered), but lasting up to months, are sometimes associated with as much deformation as a magnitude-7 regular earthquake.

In this issue, Ide *et al.* (page 76)³ present the first compilation of these slow processes' fundamental seismic parameters, and obtain relationships between them that are very different from those of regular earthquakes. These investigations are of more than academic interest: the slow-slip processes might signal times of enhanced probability for potentially hazardous regular earthquakes⁴, and the zone where they are active might delineate the lower boundary of the locked zone storing stress for the next megathrust earthquake⁵.

The fundamental descriptive parameters of regular earthquakes, such as size and duration, follow relationships over many orders of magnitude called scaling laws⁶. The primary measure of earthquake size, its seismic moment, is calculated by multiplying the rock rigidity, the slipping area and the amount of lateral slip, known as the fault offset. Slipping area and fault offset vary greatly, so the seismic moment itself varies over more than ten orders of magnitude.

Another crucial parameter, the velocity of a rupture's propagation, is quasi-constant: it is controlled dynamically by the seismic waves generated by the already broken and slipping portion of the fault, which travel at a constant velocity set by the rock's material properties. Over a wide range of sizes, therefore, the fault length involved in an earthquake is proportional to the earthquake's duration. Furthermore, a rough proportionality between fault offset and fault length leads to a well-known scaling law: earthquake duration is proportional to the cube-root of seismic moment^{6,7}.

For slow-slip events, on the other hand, the control by dynamic waves is absent. The relations between seismic moment, length, rupture velocity, duration and offset are probably therefore also different. The subdued signals, and the events' gradual beginnings and endings, make determining these quantities a tricky undertaking.

Ide and colleagues' insight³ was to recognize that many different slow-slip phenomena might be aspects of the same process, and so follow a single scaling law. This was not immediately obvious, because these processes occur on such highly varying scales of space and time, and so had been detected and analysed with different instrumentation and techniques. Most slow-slip events occur in subduction zones, but some instances have been reported⁸ on the San Andreas Fault, a very different strike-slip fault where two plates slide past each other laterally. Improved geodetic and strain monitoring since the 1990s have detected single

silent earthquakes with magnitude up to 7.5, slow aseismic afterslip following large earthquakes, and episodic slow-slip events. Improved analysis applied to more-complete archives of seismic data has also revealed long-duration, low-amplitude tremor similar to that seen below active volcanoes, and many earthquakes of anomalously low frequency.

All these slow events are distinguished from regular earthquakes of the same size, by definition, by their longer durations. But strikingly, Ide *et al.* show that, for the whole menagerie, duration is roughly proportional to seismic moment — rather than to its cube-root, as for regular earthquakes. The implication, that tremor amplitude does not grow much with increasing seismic moment or event duration, is consistent with the amplitude-limited appearance of some tremor episodes.

The authors explore the implications of their new scaling law, assuming two possible relations for how fault offset grows with fault length, L . First, they assume that the two are proportional, as for regular earthquakes. In this case, duration is proportional to L^3 , the drop in stress during an event is constant and about 100 times lower than for regular earthquakes, and rupture velocity must change as L^{-2} . The authors' alternative assumption is that fault offset is quasi-constant, limited by the accumulated strain due to plate motion since the previous major slow-slip event. This leads to a duration proportional to L^2 , a larger drop in stress for smaller events, and a rupture velocity that changes as just L^{-1} .

The data collected so far are insufficient to distinguish between the two models. The authors favour the second, pointing out that duration proportional to L^2 has the form of a diffusive process, and that fluids have been detected in the Nankai subduction zone in Japan. The fluid diffusivity that would be inferred from some of their results, however, is more than 1,000 times larger than known values in shallow tectonic crust.

A key to further progress will be more accurate location of slow events. Recent results in Japan by some members of the same research team show that the low-frequency earthquakes lie on a dipping plane at, or very near, the subduction interface⁹. But non-volcanic tremor has also been observed to the west, far above the subducting slab¹⁰, and tremor episodes in the Cascadia subduction zone off the US Pacific coast, although more difficult to locate accurately, seem to span a large range of depths¹¹. The debate on the nature of the tremor process is currently tipping towards a shear-slip origin, rather than fluid movement¹². A better idea of the space-time development of the slow phenomena would also help to adjudicate.

The approach taken by Ide *et al.*³ is a natural first step, but subject to possible pitfalls. Their lumping together of seemingly disparate events might conflate a single event with a series of events. Comparing small to large slow-slip phenomena involves comparing individual

migrating low-frequency events to parameters derived from their overall migration pattern¹³, somewhat analogous to comparing a mainshock to subevents within the mainshock. It's not entirely clear that a valid scaling law can be gleaned from comparing apples and oranges — or rather, papayas and papaya seeds. Also questionable is the authors' assumption in their analysis that slow events have a circular geometry. Considering the tectonic setting of slow slip along the lower edge of the locked portion between the overriding and downgoing plates in a subduction zone, an assumption of constant or limited width, at least for the larger events, might be more appropriate.

As a provocative first synthesis, the results of Ide *et al.* raise further questions. How long is a particular location in a large slow-slip episode active? Do events repeat in detail? Are there events intermediate between slow and regular earthquakes? That so many questions naturally arise and are already being attacked with such alacrity is a sign of the intense interest in the nature of such phenomena in the geophysical and wider community. ■

Heidi Houston and John E. Vidale are in the Department of Earth and Space Sciences, University of Washington, 4000 15th Avenue NE, Seattle, Washington 98195-1310, USA. e-mail: heidi.houston@gmail.com

1. Rogers, G. & Dragert, H. *Science* **300**, 1942–1943 (2003).
2. Ito, Y., Obara, K., Shiomi, K., Sekine, S. & Hirose, H. *Science* **315**, 503–506 (2007).
3. Ide, S., Beroza, G. C., Shelly, D. R. & Uchide, T. *Nature* **447**, 76–79 (2007).
4. Mazzotti, S. & Adams, J. *Bull. Seismol. Soc. Am.* **94**, 1954–1959 (2004).
5. Dragert, H., Wang, K. & Rogers, G. G. *Earth Planets Space* **56**, 1143–1150 (2004).
6. Kanamori, H. & Anderson, D. L. *Bull. Seismol. Soc. Am.* **65**, 1073–1095 (1975).
7. Houston, H. *J. Geophys. Res. Solid Earth* **106**, 11137–11150 (2001).
8. Nadeau, R. M. & Dolenc, D. *Science* **307**, 389 (2005).
9. Shelly, D. R., Beroza, G. C., Ide, S. & Nakamura, S. *Nature* **442**, 188–191 (2006).
10. Ohmi, S., Hirose, I. & Mori, J. *Earth Planets Space* **56**, 1185–1189 (2004).
11. Kao, H. *et al.* *Nature* **436**, 841–844 (2005).
12. Ide, S., Shelly, D. R. & Beroza, G. C. *Geophys. Res. Lett.* **34**, L03308; doi:10.1029/2006GL028890 (2007).
13. Shelly, D. R., Beroza, G. C. & Ide, S. *Nature* **446**, 305–307 (2007).

MATERIALS SCIENCE

Nanotubes see the light

Dirk M. Guldi

When it comes to having their conduction properties tweaked, carbon nanotubes are bothersome customers. One way to do it is to incorporate a photosensitive dye into the nanotubes' walls.

One of the great challenges of nanotechnology today is scaling electronic and mechanical devices down to atomic dimensions. A notably successful example has been the field-effect transistor (FET), in which a weak electric field is used to switch on and off a flow of electricity in a nearby semiconductor material. FETs are now used in their millions to, for example, amplify wireless signals. Writing in *Physical Review Letters*, Simmons *et al.*¹ report a further refinement of such nanoscale devices: the incorporation of a photosensitive dye into a single-walled carbon nanotube, so that its electrical conductivity can be controlled by light.

Single-walled carbon nanotubes — long, thin carbon 'wires' just a nanometre or so across, but up to many thousands of times longer — have exciting mechanical, optical and electrical properties that would seem to make them ideal nanoscale materials². But despite this great promise, materials scientists have encountered huge problems in actually working with nanotubes. This is partly because all methods used to synthesize them in reasonable quantities produce mixtures of metallic and semiconducting species. In addition, very strong attractive potentials build up between individual nanotubes, leading

them to stick together in ropes or bundles.

But transistor applications, for instance, require semiconducting nanotubes, and these will be overpowered by their metallic counterparts when stuck together. So far, there is no practical way, beyond individual manipulation of the nanotubes, to separate the two species, or to arrange them into molecular transistor circuits.

Some of these limitations can be overcome through the controlled covalent functionalization of the nanotubes' side walls^{3,4}: in other words, tweaking their properties by attaching covalently linked molecular 'handles'. The formation of the covalent linkages guarantees the structural integrity of the nanotube skeleton, but it also fundamentally changes the electronic structure of the individual carbon atoms. The most notable effect is that the inherent conductivity of the nanotubes is destroyed altogether.

Simmons and colleagues¹ demonstrate an alternative strategy that bypasses this problem by facilitating the integration of a molecular handle — a photosensitive dye — while on the other hand preserving the carbon atoms' original electron-orbital structure. They achieve this through a strategy of non-covalent, 'supramolecular' functionalization. Supramolecular

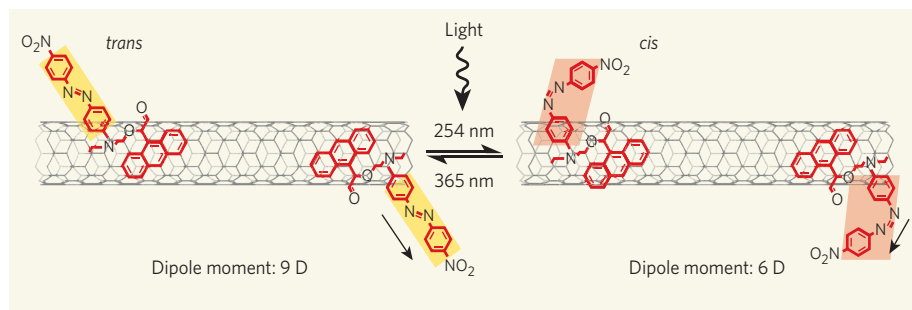


Figure 1 | Light bending and stretching. Simmons *et al.*¹ make the conductivity of carbon nanotubes responsive to light by adding molecules of the azo-based Disperse Red 1 dye to the nanotube walls. These dye molecules undergo photoisomerization, with their molecular conformation shifting around their central nitrogen double bond: from the *trans* to the *cis* form under ultraviolet light of wavelength 254 nm, and back again under blue light of 365 nm. The changes cause significant, reversible shifts in the molecules' electrical dipole moments (unit: debye, D; arrows indicate direction), and thus in the electrical conductance of a nanotube transistor as a whole.

interactions involve, for example, hydrophobic, van der Waals and electrostatic forces, and are efficient when used at short range between molecular building-blocks, but are too weak to cause intermolecular changes or form actual bonds.

Such interactions implicitly require the physical adsorption of suitable molecules onto the side walls of the nanotubes. The non-covalent immobilization of numerous polycyclic aromatic species — most notably, conjugated polymers² and small molecules such as pyrene^{6,7} — onto a nanotube surface has been at the forefront of research so far. Such additions have emerged as versatile building-blocks that can be used to modify nanotubes' solubility, fine-tune their electronic properties and exfoliate individual nanotubes from the initial bundles.

Simmons and colleagues' functionalizing addition is the commercially available Disperse Red 1 dye. Like many other azo-based dyes (that is, dyes centred around a nitrogen double bond), Disperse Red 1 undergoes a highly reversible '*cis-trans*' molecular reconfiguration, or isomerization, when exposed to light of different wavelengths. This photoisomerization involves the rearrangement of the dye's outlying molecular groups around the double bond, which cannot itself rotate. Ultraviolet light of wavelength 254 nm initiates the conversion from the *trans* to the *cis* form, whereas blue light of 365 nm brings about the back-conversion from *cis* to *trans* (Fig. 1).

Importantly, the reconfiguration of the dye molecule causes a significant change in the electrical dipole moment along its principal axis. Simmons *et al.* document impressively that this shift also modifies the local electrostatic potential in the nanotube and, in turn, modulates its conductance by shifting the threshold voltage at which current flows. With a dye coverage of 1–2%, the photoisomerization induced a shift in the threshold voltage of up to 1.2 V. In particular, a carbon-nanotube transistor exhibiting p-type behaviour (that is, the flow of holes, carriers of positive electric charge) saw its threshold voltage increase

by 1 V to higher, positive gate voltages.

As far as integrating single-walled carbon nanotubes into transistors is concerned, the technology described by Simmons and colleagues² has several advantages. First, the functionalized transistors show repeatable switching for many cycles and a modest switching time of around 2 seconds. The authors believe that decreasing the dye concentration could further accelerate the switching. Second, the low light intensities (about $100 \mu\text{W cm}^{-2}$) needed to modulate the functionalized transistors are in stark contrast to those required to induce intrinsic nanotube photoconductivity (typically 1 kW cm^{-2}). Finally, the authors' synthesis technique is versatile enough to immobilize different dyes onto the nanotubes. That should allow the tuning both of the wavelengths that initiate the photoisomerization and the magnitude of the switching, independently of the nanotube's electronic structure.

In that light, the work is surely a breakthrough in implementing carbon nanotubes in optoelectronic devices — technologies from photovoltaic cells to flat-panel displays that have a bright future.

Dirk M. Guldi is at the Institut für Physikalische Chemie I, Friedrich-Alexander-Universität Erlangen-Nürnberg, Egerlandstraße 3, 91058 Erlangen, Germany.
e-mail: dirk.guldi@chemie.uni-erlangen.de

1. Simmons, J. M. *et al.* *Phys. Rev. Lett.* **98**, 086802 (2007).
2. Reich, S., Thomsen, C. & Maultzsch, J. *Carbon Nanotubes: Basic Concepts and Physical Properties* (Wiley-VCH, Weinheim, 2004).
3. Georgakilas, V. *et al.* *J. Am. Chem. Soc.* **124**, 760–761 (2002).
4. Hirsch, A. *Angew. Chem. Int. Edn* **41**, 1853–1859 (2002).
5. Star, A. *et al.* *Angew. Chem. Int. Edn* **40**, 1721–1725 (2001).
6. Chen, R. J., Zhang, Y., Wang, D. & Dai, H. *J. Am. Chem. Soc.* **123**, 3838–3839 (2001).
7. Ehli, C. *et al.* *J. Am. Chem. Soc.* **128**, 11222–11231 (2006).

Correction

In the News & Views article "Computing: The security of knowing nothing" by Bernard Chazelle (*Nature* **446**, 992–993; 2007), an error crept into reference 2. The correct reference is Barak, B. & Sahai, A. *Proc. 46th IEEE Symp. Found. Comput. Sci.* 543–552 (2005).



50 YEARS AGO

A note released recently by the Rockefeller Institute in New York describes a small radio-transmitting capsule which can be swallowed like a medicinal pill and, as it passes through the body, signals the activity of the digestive tract. This 'radio pill', as it is termed...comprises a plastic capsule, $1\frac{1}{8}$ in. long and $\frac{4}{10}$ in. in diameter, containing a tiny transistor oscillator... In one end of the capsule is housed a minute replaceable storage battery which supplies the small amount of electrical power required and has a life of fifteen hours... When the 'pill' is swallowed by a patient and pursues its course through the gastrointestinal tract, it transmits signals the frequency of which is varied by the pressure changes in the tract. These signals from the capsule are picked up on a suitable receiver with an antenna held close to the body.

From *Nature* 4 May 1957.

100 YEARS AGO

Dr H. Charlton Bastian re-affirms his conviction that living organisms continue to arise from not-living material... In the first part of his book he points out that inorganic evolution (recently studied in ways not a little upsetting) has not stopped, and argues against the dogmatism of those who, while admitting that archeobiosis probably occurred very long ago, refuse to discuss the possibility of its occurrence now. Because it has been shown that maggots are not really produced by the flesh in which they crawl, it does not follow that minute specks of living matter may not arise *de novo* in suitable not-living fluids... In a fluid believed to be quite not-living, minute living creatures appear, but observation cannot decide whether they arise from invisible germs or pre-existing organisms, or "whether they have come into being in the mother liquid as a result of life-giving synthetic processes."

From *Nature* 2 May 1907.

50 & 100 YEARS AGO

OBITUARY

F. Clark Howell (1925–2007)

Palaeoanthropologist who defined a discipline.

F. Clark Howell was the principal architect and prime mover of the multidisciplinary study of human evolution in the past half-century. In a career during which he came to personify palaeoanthropology, Howell distinguished himself through his unbounded curiosity, unparalleled scholarship, inspirational leadership, modesty and humanity. He died in Berkeley, California, on 10 March.

Howell was born in Kansas City in 1925. Raised on a small farm, he attended a one-room Kansas schoolhouse until the 1937 recession changed much of the US Midwest to a dust bowl, and Howell's father became a travelling salesman.

After service as a navy signalman in the Pacific during the Second World War, Howell met the palaeontologists George Gaylord Simpson, Franz Weidenreich and Ralph von Koenigswald, then working at the American Museum of Natural History in New York. These meetings proved Howell's inspiration, and he enrolled under the GI Bill of Rights as an undergraduate at the University of Chicago in 1947, completing a PhD in anthropology there in 1953.

Howell went on to teach anatomy at Washington University, St Louis, and in 1954 embarked on his first journey to a mostly still-colonial Africa. During this six-month study tour, Howell became friends with such pioneers of African prehistory as Raymond Dart, and Louis and Mary Leakey. Returning to the University of Chicago as assistant professor in 1955, Howell, together with Sherwood Washburn, helped Africa to emerge as the focus of a new, multifaceted approach to research on human origins and evolution that brought together strands from geology, climatology, ecology, archaeology, evolutionary biology and many other disciplines.

In 1957 and 1958, together with his wife Betty, Howell excavated 'Acheulean' sites containing hand-axes from the Middle Pleistocene at Isimila in the highlands of what was then Tanganyika. In 1959, he made a preliminary survey of fossil-bearing deposits in the lower Omo Valley of southern Ethiopia, and was in Nairobi, Kenya, to receive at first hand the news of Mary Leakey's momentous discovery of the early hominid *Australopithecus* in the Olduvai Gorge.

From the very start of his career, however, Howell's interest in human evolution was not confined to Africa, but spanned all continents and epochs. From the 1950s onwards, he published work bringing together Eurasian palaeobiology, archaeology and geology.

In 1956, he attended the conference marking the centennial anniversary of the Neanderthal fossil discovery in Germany, and last summer was the keynote speaker at the 150th anniversary celebrations in Bonn.

Howell displayed an unparalleled ability to operate effectively at the interface of numerous disciplines. And more than any other biological anthropologist, he was able to maintain a career-long balance that met the demands of substantive and extended fieldwork in the Old World; of professional organization, communication and teaching; and of a commitment to promoting and funding research by young scholars from all over the world.

Although Howell conducted three excavation seasons at Acheulean sites in Spain in the 1960s, his signature work came with his leadership of the American contingent of the International Omo Research Expedition in southern Ethiopia between 1967 and 1973. This project put his philosophy and skills on international display, and set standards for the field as we now know it. Thousands of vertebrate fossils, including dozens of hominid specimens, were recovered from precisely documented stratigraphic and palaeoenvironmental contexts. The era of multidisciplinary, international palaeoanthropology had begun.

Howell worked tirelessly to make the knowledge generated by his science accessible to a general audience, and so attracted new generations of students to it. In the 1960s, with illustrator Jay Matternes and author Maitland Edey, he created the *Early Man* volume for the Time-Life *Nature* series. He was senior scientific adviser to MGM for an Emmy-nominated network-television documentary, *The Man-Hunters*, that featured his Omo fieldwork.

Howell left Chicago in 1970 to join the University of California, Berkeley, founding the Laboratory for Human Evolutionary Studies (now the Human Evolution Research Center), and continuing his multidisciplinary researches there. Berkeley was thus placed at the forefront of human-origins research. On the international front, hard on the heels of US President Richard Nixon's historic trip to the People's Republic of China, Howell led the first National Academy of Sciences anthropological delegation to that country in 1975.

In the early 1980s, Howell returned to Spain for further excavations, moving on at the decade's end to extensive surveys



and excavations in Turkey. His official retirement in 1991 allowed him to become even more active in advising, writing, conducting field and laboratory work, and enjoying life. The Howells travelled abroad almost yearly to visit friends and institutions in dozens of countries across the world. Howell also read voraciously in many languages throughout his career. His knowledge was correspondingly encyclopaedic, covering subjects as diverse as Neogene carnivoran palaeontology, Turkish tectonics, classical jazz, palaeolithic typologies and molecular systematics.

In science, Howell's influence went beyond academia. His indefatigable efforts with the Leakey Foundation over 38 years enabled generations of students to expand and enrich our knowledge of human evolution. In the words of a former student, Howell "made good things happen by putting the right people together in a common cause". During his final four years, he did this again by assembling and steering the largest-ever palaeoanthropological research enterprise, the Revealing Hominid Origins Initiative funded by the National Science Foundation's HOMINID programme, aiming to illuminate the trunk of humanity's family tree.

Howell's achievements and published work, honoured by numerous awards and appointments, were broad, and will be lastingly influential. In a field where transient celebrity is sometimes associated with finding particular hominid fossils, Howell never sought popular adulation. He was a scholar's scholar, as the many tributes paid by colleagues on his death show (http://herc.berkeley.edu/fc_howell_memorial). He brought his unique combination of knowledge, wisdom, vision and generosity to bear in steadfast advocacy of the developing science of human evolution.

Tim D. White

Tim D. White is at the Human Evolution Research Center and the Department of Integrative Biology, University of California at Berkeley, Berkeley, California 94720, USA.
e-mail: timwhite@berkeley.edu

PALAEOCLIMATOLOGY

Evidence for hot early oceans?

Arising from: F. Robert & M. Chaussidon *Nature* **443**, 969–972 (2006)

The oxygen isotopes in sedimentary cherts (siliceous sediments) have been used to argue that the Precambrian oceans were hot — with temperatures of up to 70 °C at 3.3 Gyr before present¹. Robert and Chaussidon² measure silicon isotopes in cherts and arrive at a similar conclusion. We suggest here that both isotope trends may be caused by variations in seawater isotope composition, rather than in ocean temperatures. If so, then the climate of the early Earth may have been temperate, as it is today, and therefore more consistent with evidence for Precambrian glaciations and with constraints inferred from biological evolution.

By focusing on cherts, Robert and Chaussidon² have overlooked the much more extensive marine-carbonate oxygen-isotope database³. Marine carbonates can be more easily demonstrated by using strontium isotopes^{3,4}, for example, to relate to the global ocean, whereas two different mineralogies — dolomite [$\text{CaMg}(\text{CO}_3)_2$] and calcite (CaCO_3) — are studied in parallel. According to the oxygen- and silicon-isotope data of Robert and Chaussidon², surface seawater temperatures decreased by 50–60 °C during the Precambrian, mostly between 2.0 Gyr and 1.0 Gyr ago. Such a major change in ocean temperature ought to be recorded in the marine carbonate record by a rise in $\delta^{18}\text{O}$ values of at least 10‰; however, no such rise is evident.

If only the average values of each unit are plotted (to reduce bias from individual studies with many samples), then a much smaller (~5‰) rise in carbonate $\delta^{18}\text{O}$ is seen between 2.5 Gyr and 0.8 Gyr ago³. This rise could represent a much more modest temperature decrease of about 20 °C, which would be more consistent with evidence for glaciation and from fossil and biomarker records of eukaryo-

tes throughout this period in Earth's history.

On the other hand, secular changes in marine chert and carbonate $\delta^{18}\text{O}$ values, and perhaps $\delta^{30}\text{Si}$ values as well, need have nothing to do with temperature. The largest shift in carbonate $\delta^{18}\text{O}$ occurred more recently during the early Palaeozoic⁴ and has been interpreted as resulting mainly from an increase in seawater $\delta^{18}\text{O}$ (see refs 4, 5, for example). This is because of the implausibly hot global climate that would need to be inferred were temperature the sole controlling factor on carbonate $\delta^{18}\text{O}$ during this period of animal expansion.

The secular change in $\delta^{18}\text{O}$ of cherts can also be explained if seawater isotopic composition changes with time^{6,7}. This possibility, much discussed in the literature, has been overlooked by Robert and Chaussidon². If the increase in marine chert and carbonate $\delta^{18}\text{O}$ from the Archaean to the present day were caused by an equivalent increase in seawater $\delta^{18}\text{O}$, then it must have been related to a decrease in the relative influence of low- versus high-temperature alteration of the ocean crust^{5–7}. Such a decrease could have been caused by changes in tectonic styles associated with decreasing geothermal heat flow^{7,8}. In particular, if the mid-ocean ridge crests were shallower in the past, hydrothermal circulation systems may have penetrated less deeply, and sea water may have interacted with basalts at lower temperatures⁸.

Such a change in mid-ocean-ridge circulation systems may also explain the trend towards higher seawater $\delta^{30}\text{Si}$ during the Precambrian noted by Robert and Chaussidon². In their model, the silicon isotope composition of sea water is controlled by the fraction of silica that leaves the ocean in sediments, f , compared with the fraction, $1 - f$, that leaves as a result of hydrothermal silicification of oceanic crust. The

latter fraction, $1 - f$, depends on the temperature difference between the circulating hydrothermal fluid, T_{hyd} , and sea water, T_{SW} . The smaller the difference between these two numbers, $T_{\text{hyd}} - T_{\text{SW}}$, the smaller is the fraction $1 - f$ and the greater is the fraction f . The authors' data² show that the early Precambrian was characterized by a greater value of f , or a smaller value of $1 - f$, which they interpret as indicating an increase in T_{SW} . The implicit assumption here is that T_{hyd} remains constant. We propose that this change might equally well be caused by a decrease in T_{hyd} , while T_{SW} remains more or less constant. Such a model, driven by changes in heat flow and plate tectonics⁸, offers a natural explanation for the secular changes in both $\delta^{18}\text{O}$ and $\delta^{30}\text{Si}$ in sedimentary rocks — one that does not conflict with the observed record of Precambrian glaciations and biological evolution.

Graham A. Shields*, James F. Kasting†

*Geologisch-Paläontologisches Institut und Museum, Westfälische-Wilhelms Universität, Münster, 48149 Münster, Germany

†Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Received 10 November 2006; accepted 16 April 2007.

1. Knauth, L. P. & Lowe, D. R. *Geol. Soc. Am. Bull.* **115**, 566–580 (2003).
2. Robert, F. & Chaussidon, M. *Nature* **443**, 969–972 (2006).
3. Shields, G. & Veizer, J. *Geochim. Geophys. Geosys.* **3**, 1031 (2002).
4. Veizer, J. *et al. Chem. Geol.* **161**, 59–88 (1999).
5. Wallmann, K. *Geochim. Cosmochim. Acta* **65**, 2469–2485 (2001).
6. Perry, E. C. *Earth Planet. Sci. Lett.* **3**, 62–66 (1967).
7. Walker, J. C. G. & Lohmann, K. C. *Geophys. Res. Lett.* **16**, 323–326 (1989).
8. Kasting, J. F. *et al. Earth Planet. Sci. Lett.* **252**, 82–93 (2006).

Competing financial interests: declared none.
doi:10.1038/nature05830

PALAEOCLIMATOLOGY

Robert & Chaussidon reply

Replying to: G. A. Shields & J. F. Kasting *Nature* **446**, doi:10.1038/nature05830 (2007)

The possibility of a change in $\delta^{18}\text{O}$ values of the oceans has been discussed for the past 30 years¹ and the comment by Shields and Kasting² does not really bring any new insight into this issue. The carbonate $\delta^{18}\text{O}$ curve³ is indistinguishable from that of cherts and shows a huge scatter caused by local geological processes. This scatter in $\delta^{18}\text{O}$ (up to 20‰ at a given age) far exceeds the sensitivity of the isotope thermometer (about 2‰ per 10 °C). It therefore prevents any precise additional test of the model and no consensus was ever reached

in the literature on this debate from oxygen-isotope data alone. Our approach⁴, by contrast, was to try to find another isotopic proxy ($\delta^{30}\text{Si}$) that could bring new and independent constraints to test the hypothesis of hot oceans in the early Precambrian. Our interpretation of the $\delta^{30}\text{Si}$ values in terms of temperature is fully coherent not only with the $\delta^{18}\text{O}$ of cherts but also with that of marine carbonates, and certainly does not overlook these data. In addition, we indicated⁴ that our sample set has not the time resolution that would be necessary

to look for an effect on silicon isotopes of Precambrian global glaciations. No chert samples from such periods were analysed.

A change in the $\delta^{18}\text{O}$ of sea water down to –15‰ (as indicated by the difference between modern and the Archaean chert values) that would result from a drastic change with time of the hydrothermal circulation at oceanic ridges would seem to be unattainable because there is no mineral assemblage that can fractionate oxygen isotopes to the extent required by models involving alteration of the oceanic crust^{5,6}.

In addition, the $\delta^{18}\text{O}$ values of mineral pairs in cherts (phosphate–silica)⁷ used to calculate the seawater $\delta^{18}\text{O}$ independently never give a $\delta^{18}\text{O}$ in sea water that departs by more than 5‰ (in both directions) from the present-day value.

Shields and Kasting² suggest that in our model⁴ a variation of the temperature of the hydrothermal fluids circulating in the oceanic crust, T_{hyd} , could be responsible for the change in $\delta^{30}\text{Si}$, instead of a variation in the seawater temperature. In our model, T_{hyd} varies between 190 and 230 °C. This range stands for the range of temperature at which most silica precipitation occurs from a fluid injected into the crust at about 300 °C. First-order calculations show that this T_{hyd} in the Archaean should be close to 110 °C in order to reproduce the $\delta^{30}\text{Si}$ variations observed in our data, with a constant seawater temperature of 10 °C.

How is it possible to lower T_{hyd} by about

100 °C while the Precambrian crust temperature was probably higher than the present-day value? Can such low-temperature fluids efficiently leach silicon in the oceanic crust and precipitate it to produce the silicon isotopic fractionation required to explain the chert data? Is it possible to obtain an equilibrium between dissolved and precipitated silica at such a low temperature, and at what rate? Although these issues cannot be simply addressed from existing experimental data or theoretical considerations, they might nevertheless be tested in cherts with the ion microprobe by quantifying the isotope homogeneity at the micrometre scale. Oxygen and silicon isotope studies of hydrothermal silica (silica veins in Precambrian rocks) should constrain the temperature of hydrothermal reactions and the isotopic composition of the fluids.

François Robert* & **Marc Chaussidon†**

*Muséum National d'Histoire Naturelle, Centre National de la Recherche Scientifique Laboratoire d'Etude de la Matière Extraterrestre Nanoanalyses, UMS 2679, 75005 Paris, France

†Centre de Recherches Pétrographiques et Géochimiques, Centre National de la Recherche Scientifique, BP20, 54501 Vandœuvre-Lès-Nancy, France

1. Perry, E. C. & Tan, F. C. *Geol. Soc. Am. Bull.* **83**, 647–664 (1972).
2. Shields, G. A. & Kasting, J. F. *Nature* **446**, doi: 10.1038/nature05830 (2007).
3. Kasting, J. F. *et al. Earth Planet. Sci. Lett.* **252**, 82–93 (2006).
4. Robert, F. & Chaussidon, M. *Nature* **443**, 969–972 (2006).
5. Lécuyer, C. & Allemand, P. *Geochim. Cosmochim. Acta*, **63**, 351–361 (1999).
6. Simon, L. & Lécuyer, C. *Geochim. Geophys. Geosys.* **6**, 1–10 (2005).
7. Karhu, J. & Epstein, S. *Geochim. Cosmochim. Acta*, **50**, 1745–1756 (1986).

doi:10.1038/nature05831

An integrated model of kimberlite ascent and eruption

Lionel Wilson¹ & James W. Head III²

Diatremes are carrot-shaped bodies forming the upper parts of very deep magmatic intrusions of kimberlite rock. These unusual, enigmatic and complex features are famous as the source of diamonds. Here we present a new model of kimberlite ascent and eruption, emphasizing the extremely unsteady nature of this process to resolve many of the seemingly contradictory characteristics of kimberlites and diatremes. Dyke initiation in a deep CO₂-rich source region in the mantle leads to rapid propagation of the dyke tip, below which CO₂ fluid collects, with a zone of magmatic foam beneath. When the tip breaks the surface of the ground, gas release causes a depressurization wave to travel into the magma. This wave implodes the dyke walls, fragments the magma, and creates a 'ringing' fluidization wave. Together, these processes form the diatreme. Catastrophic magma chilling seals the dyke. No precursor to the eruption is felt at the surface and the processes are complete in about an hour.

From a geophysical standpoint, kimberlites are fluid-rich materials, ultramafic in overall composition, that extract xenoliths rapidly and erupt explosively from great depths. The presence of diamonds in some kimberlites indicates a source at pressures of at least 6–8 GPa, the pressure at which diamond is stable relative to graphite at mantle temperatures; this corresponds to depths of ~200–250 km. From a petrologic standpoint, kimberlite is a volatile-rich, very-low-silica igneous rock. Mineral deposit geologists view kimberlite as a porphyritic alkalic peridotite with abundant phenocrysts of olivine and phlogopite. Finally, from a volcanologic standpoint, diatremes are funnel-shaped breccia pipes that extend to as much as 2,500 m in depth, and are thought to form¹ by “hydrovolcanic fragmentation and wall rock collapse ...[and]...may underlie maars and grade at depth into dykes.” This range of definitions illustrates the diverse nature and unusual properties of diatremes and kimberlites², and suggests why no single model currently exists that successfully explains their often contradictory characteristics. We summarize these characteristics to provide a basis for developing a comprehensive model for the ascent and eruption of kimberlites and the formation of diatremes and associated features.

Background and characteristics

Early studies of kimberlites³ showed that they occur both as carrot-shaped vertical intrusions (pipes or diatremes) and as tabular dykes known as fissure kimberlites, but their connections were not fully appreciated until the early 1970s when the classic analyses of Dawson⁴ and Hawthorne⁵ established the basic principles of kimberlite magmatism. These authors recognized the existence of hot mobile kimberlite magmas capable of undergoing differentiation, the occurrence of pyroclastic and epiclastic kimberlites, and the gradation with increasing depth of diatremes into non-brecciated hypabyssal kimberlites³. From this time on, kimberlites were recognized as volatile-rich ultrabasic magmas, the evolution and emplacement of which could be described in terms of standard processes of differentiation, intrusion and extrusion. Diatremes thus became only one particular manifestation of a more general magmatic style³. This led to the understanding of the relationships between the major components of a kimberlite magmatic system (Fig. 1), which include crater, diatreme

and hypabyssal facies and three textural genetic groups of rocks, each associated with a particular style of magmatic activity (see, for example, refs 3–6). These components are described below.

Crater facies kimberlites include pyroclastic rocks and epiclastic rocks. Kimberlitic magmas rarely produce lava flows but typically form pyroclastics that, where studied in detail⁷, display several types of deposits. From oldest (lowest) to youngest, these can include basal breccias, poorly stratified coarse pyroclastic deposits (tuffs and tuff breccias) containing fragments of kimberlite, country rock and mantle-derived xenoliths, well-stratified tuffs (alternating layers of coarse lapilli-sized tuffs and laminae of finer ash-sized tuffs) and, finally, epiclastic lacustrine deposits. In many cases, graded beds and depositional features seem to be absent, leading to the interpretation that such tuffs are primarily airfall⁷. Fluvial reworking of tuffs in crater lakes produces epiclastic kimberlites. Volumes of pyroclastic deposits are small and they are typically confined to craters and to thinly bedded tuff-rings; subsequent shallow magmatic intrusions or extrusive lava flows do not typically follow pyroclastic eruptions. Erosion usually follows quickly, but marginal slumping and downfaulting may preserve crater rim facies. These deposits have some similarities to hydrovolcanic tuff rings⁸.

Diatreme facies kimberlites underlie crater facies kimberlites (Fig. 1). They occur in carrot-shaped bodies with circular to elliptical cross-sectional areas that have vertical axes and steeply dipping margins that converge and terminate at depth in a root zone, where the diatreme expands, contracts, or splits up into an irregularly shaped multiphase intrusion of hypabyssal kimberlite³. The commonest rocks in the diatreme facies are tuffisitic kimberlite breccias, containing pelletal lapilli, abundant angular to rounded country rock inclusions (mostly a few centimetres down to microscopic), and discrete and fractured grains of olivine, garnet and ilmenite mega- and macrocrysts, set in a fine-grained matrix of microcrystalline diopside and serpentine. The matrix quickly undergoes alteration and replacement by clays and secondary calcite³. Typically, one to three texturally distinctive varieties of tuffisitic kimberlite breccias are seen in diatreme zones.

Hypabyssal facies kimberlites (Fig. 1) are rocks formed by the crystallization of volatile-rich kimberlitic magma and exhibit igneous

¹Environmental Science Department, Lancaster University, Lancaster LA1 4YQ, UK. ²Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA.

textures and effects of magmatic differentiation; often they contain sufficient country rock xenoliths to be called kimberlitic breccias. These occur as dykes and form the root zones of diatremes (Fig. 1). Kimberlitic dykes are typically vertically dipping with 1–3 m widths, but can be up to 10 m wide³, and commonly form swarms of parallel features. Most dykes are single intrusions and pinch out towards the surface, thickening with depth; many show evidence of flow differentiation, glassy selvages are absent, and contact metamorphic effects are slight. Some dykes are observed to expand along strike into lenticular features 10–20 times the dyke width and up to 100 m in length; these are termed ‘blows’ and may represent the lowermost portions of root zone intrusions³. Internal dykes are common in most diatremes and root zones, but are small, rootless, sinuous and pinch out laterally and vertically, cross-cutting intrusions within the pipes but not extending into the surrounding country rock. Most have no preferred orientation and may be localized at the dyke–wall-rock contact or at the contact between discrete intra-diatreme intrusions³. Subsequent or cross-cutting dykes are extremely rare, suggesting that the diatreme-forming event is the closing stage of kimberlitic magmatism. Kimberlitic sills are relatively rare and plutonic kimberlitic complexes are unknown³.

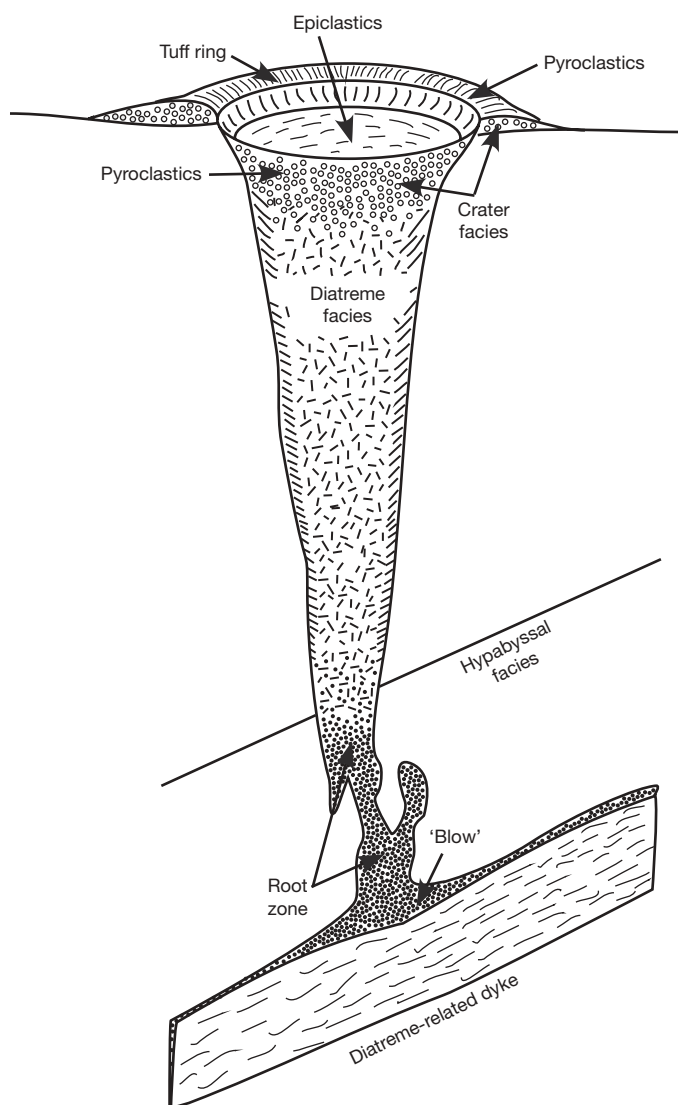


Figure 1 | Model of an idealized kimberlite magmatic system illustrating the relationships between crater, diatreme, and hypabyssal zones and facies rocks. Not to scale. Modified from Fig. 3.1 on page 30 of ref. 3, with kind permission of Springer Science and Business Media.

Summary of distinctive characteristics

The major characteristics that need to be explained in any model for the formation of kimberlite diatremes are summarized in Table 1. Relative to normal magmatic eruptions, the most unusual of these characteristics raise the questions of how is it possible to: (1) transport diamonds from the mantle (where they are stable) to the surface (where they are metastable) fast enough to avoid significant alteration over the intervening range of depths at which they are unstable; (2) transport significant quantities of mantle xenoliths to near-surface levels; (3) produce intrusive pyroclastic kimberlitic material at depths where pressures would normally preclude magma fragmentation by volatile expansion, and then separate most of the volatiles from the pyroclastic materials; (4) generate extensive fracturing and brecciation of host rocks in many parts of the system, with minimal contact metamorphic effects; and (5) produce complex shallow structures (diatremes and root zones) in which an unusual widening of the conduit system occurs, sometimes with extensions that do not connect to the surface, and where complex mixtures of fragmented material and coherent dyke segments are present.

Models for kimberlite emplacement and diatreme

Attempts to explain the characteristics outlined above have included explosive liberation of highly compressed magmatic gases at shallow depths to produce a volcanic vent (diatreme)⁹ and small, subsurface, discrete and repeated upwardly migrating explosions causing explosive boring¹⁰. Fluidization—the transportation and circulation of particles by high-velocity gas flow¹¹—has been extensively applied to the formation of diatremes, with some^{4,12–15} postulating that a gas-charged kimberlite magma breaks through to the surface, and that adiabatic expansion of magmatic gas then enlarges the explosion vent, which is subsequently filled with fluidized fragmented kimberlite that drills upward with a sandblasting effect. On the basis of the similarities of maars/tuff cones to the surface expressions of diatremes, several authors have strongly advocated that diatremes are caused by hydrovolcanic activity, the interaction of magma and near-surface water (for example, refs 16–18): magma rising in pre-existing cracks encounters water-rich zones and produces hydrovolcanic explosions that fragment and chill magma while simultaneously brecciating the country rock. Continued and repetitive activity results in a diatreme, the conduit surrounded by brecciated and spalled rock.

Table 1 | Key characteristics of type 1 kimberlite diatremes

Structure and shape
Distinctive carrot shape of diatreme
Three distinctive regions and facies (crater, diatreme and hypabyssal)
Characteristics
Kimberlitic mantle-derived composition and mineralogy
Volatile-rich nature of kimberlite magma, dominated by CO ₂
Angular clasts of shallow country rock, typically devoid of kimberlitic coating
Pronounced sphericity of lapilli (mostly 1–10 mm in size)
Glassy or microcrystalline spherules, implying rapid quenching
Smaller amounts of more rounded lower crustal and mantle material, exposed deeper within the diatreme
Olivine-cored pelletal lapilli surrounded by usually altered quenched kimberlitic melt or glass
Features usually lacking
Extensive extrusive deposits
Thermal metamorphism
Exposed plutonic complexes
Subsequent dykes
Vesicles and composite lapilli
Implications for emplacement models
Apparent rapid emplacement, preventing destruction of diamonds in transit
Implied low temperatures of emplacement, despite magmatic transport

The difficulty of having one single mechanism account for the diversity of characteristics of kimberlites and diatremes led Clement^{19,20} to propose more complex models combining different processes (embryonic pipes modified into diatremes by fluidization and fluctuation in magma flow rates, producing changes in pressure and implosion and shattering of country rocks). Mitchell³ suggested that no single process could be responsible for the formation of the complex structure of kimberlites and diatremes and that root zone processes may combine with hydrovolcanic processes at shallower depths to produce the observed features³. Using new insights from the fluid mechanics of dyke propagation, we propose a model that accounts for most of the features summarized in Table 1.

A new model for ascent and eruption

We propose that essentially all of the rise of kimberlitic magma to the surface takes place via rapid propagation of a dyke from unusually great depths (Fig. 2) rather than the much slower propagation of a diapir to shallower depths before transitioning to a dyke. This very deep dyke initiation and propagation minimizes thermodynamic problems associated with transporting diamonds from mantle depths to the surface. Dyke propagation is initiated in a diapirically rising^{21,22} CO₂-rich mantle source region^{23,24} when the percentage of partial melt in the source exceeds the critical level allowing upward drainage to begin at a rate so large that the maximum strain rate at which the matrix can deform plastically is exceeded^{25,26}. Previous kimberlite dyke propagation analyses have emphasized magma flow once a dyke has reached the surface. Our new model focuses on the pressure distribution in a dyke during its ascent.

Stage 1. This stage involves dyke tip propagation out of a deep source region and CO₂ fluid segregation. The dyke sets out from the mantle source^{27,28} at a depth of ~250 km, where the pressure is ~8 GPa (Fig. 2a). The magma in the dyke could contain as much as 20 wt% of CO₂ (ref. 29), and the pressure-dependent solubility of this volatile is such that the dyke tip pressure, in attempting to reach the lowest possible value in order to maximize the magma flow speed, would initially be buffered at ~2 GPa by the release of 90% of the available CO₂ (ref. 25) as a supercritical fluid filling the cavity behind the dyke tip. The difficulty of diffusing the volatile phase into the cavity causes a foam layer to rapidly develop beneath the tip cavity (Fig. 2b). Fluid bubbles in the foam burst into the cavity, and the cavity pressure decreases³⁰ to the value at which the bubble volume fraction is ~0.7–0.8, the condition under which³¹ volcanic foams commonly disintegrate into a continuous volatile phase containing entrained droplets of the magmatic liquid. The CO₂ density at which ~20 wt% forms bubbles occupying ~75% of the volume at magmatic temperatures is ~220 kg m⁻³. Using thermodynamic data for supercritical CO₂, this corresponds to a pressure of ~70 MPa. Thus, the pressure decreases upward from ~2 GPa to ~70 MPa across the foam layer. The supply of CO₂ is constantly renewed by streaming of degassed magma to the sides of the dyke as the dyke grows both upward and laterally, exposing fresh undegassed magma in the dyke centre. The tip cavity grows in length in a manner controlled by the elastic stresses acting over the upper part of the dyke³² and may be 2–4 km long by the time the dyke tip is near the surface.

Stage 2. This stage involves dyke ascent and wall fracturing. As magma ascends from the source region at a pressure of ~8 GPa to the base of the foam layer at ~2 GPa it will cool adiabatically from ~1,650 K to ~1,450 K (ref. 33). Its passage through the foam layer from a pressure of ~2 GPa to ~70 MPa can be treated as the adiabatic expansion of a pseudo-gas³⁴, resulting in cooling from ~1,450 K to ~1,110 K. The pressure difference in excess of the static weight of the magma column (source pressure minus tip pressure) will be fixed at ~8 GPa, and on subtracting the static weight of the magma column the pressure difference driving the magma motion is found to vary from ~8 GPa at great depth to ~1 GPa near the surface. Dividing by the length of the magma column yields a pressure gradient ranging from ~1 MPa m⁻¹ at great depth to ~4 kPa m⁻¹ near the surface. The average pressure

gradient of more than 60 kPa m⁻¹ is ~20 times larger than the gradients driving basaltic eruptions from shallow magma reservoirs³⁵ and leads to the magma flowing upward in a turbulent manner with an average rise speed of ~30–50 m s⁻¹, implying an extremely short transit time of only about 1 hour (ref. 36; Fig. 2c).

The dyke tip pressure is buffered at ~70 MPa during the entire rise to the surface but the external pressure decreases owing to the decreasing overburden. Over most of the rise distance of the dyke, wall rocks adjacent to the propagating dyke tip region are fractured in tension (Fig. 2c). Country rock is torn from the walls to become xenoliths, and quickly sinks through the CO₂ fluid to become engulfed by and incorporated into the underlying magmatic foam. Over the last several kilometres of magma rise, the stress across the dyke walls changes from tensile to compressive, with the possibility of intrusion of small dykelets owing to the decreasing overburden pressure. The relative abundance of xenoliths produced by wall rock fracturing will be a function of rock strength and position in the crust relative to the evolving differential stress; the most important factor will be the elapsed time over which the wall rocks are exposed to dyke emplacement, favouring deep xenoliths.

Stage 3. Next, the dyke tip breaks the surface, vents CO₂ gas and implodes the walls. The propagating dyke is convex upward along its length and first reaches the surface at its highest central point, immediately starting to vent the CO₂ fluid from the dyke tip cavity (Fig. 2d). On being exposed to the surface, the supercritical fluid expands adiabatically to atmospheric pressure to become a subcritical gas, producing an initial vulcanian explosion. Its upward velocity increases from the ~20 m s⁻¹ rise speed of the dyke to a speed that depends on the amount of clastic material which it carries. If a pure gas is released its speed will be ~1.4 km s⁻¹ and it will have cooled adiabatically to ~300 K. If the gas is loaded with all of the magma in the form of small droplets its speed will be ~600 m s⁻¹ and its temperature ~680 K, and if it is loaded by magma droplets and an equal mass of host rock wall fragments, then the values will be ~300 m s⁻¹ and ~500 K.

The expansion wave causing this violent acceleration of the cavity fluid will propagate downward through it at about half the speed of sound in CO₂ at the magmatic temperature (that is, ~300 m s⁻¹) emptying a typically 3-km-deep cavity in ~10 s. The mixture of gas and entrained particles ejected into the atmosphere will produce a classic Prandtl jet that will begin to interact with the surrounding air to produce a plinian or sub-plinian eruption plume (Fig. 2e). The dyke will rapidly centralize along the widest portion (almost certainly the central part that reaches the surface first and within which the rise speed of magma and gas is greatest) and the remainder of the upper part of the gas-filled dyke will rapidly close. This produces a linear fractured and crushed zone with little to no evidence of associated magma, in the centre of which is a much less elongate central vent. Within this vent, the break-through and the ensuing gas jet will rip wall rock from the uppermost country rock, and the proportion of shallow country rock should be high in the initial ejecta deposits around the vent. Simultaneously, the sharp decrease in pressure caused by the gas venting will fracture and implode the walls of the upper part of the dyke.

Stage 4. The depressurization wave initiated by the gas venting next propagates down through the layer of magmatic foam at about half the speed of sound in the foam (that is, ~50 m s⁻¹; ref. 37), expanding the bubbles and disrupting the foam into magma droplets and released gas (Fig. 2e). The wave continues into the underlying magma at about half the speed of sound in the bubble-free liquid (that is, ~800 m s⁻¹), and more CO₂ is released in the magma-filled portion of the dyke, forming additional foam which also expands and is disrupted (Fig. 2f). In both cases the assemblage of gas and entrained magma droplets will cool from ~1,110 K to ~680 K as long as the front of the expansion wave is maintained at atmospheric pressure. During the disruption and expansion process (Fig. 2f), surface tension will form the resulting liquid fragments into spheres to optimize the surface area to volume ratio. These spheres would incorporate any solid particles in the rising magma (olivine

phenocrysts, xenolithic grains) that had acted as nuclei for the volatile bubbles, and the very rapid adiabatic cooling of these particles would produce glassy or microcrystalline spherules, cored by phenocrysts and xenoliths, the pelletal lapilli of the tuffisitic kimberlite breccias. Cooling would be so rapid (going from magmatic to room temperature in seconds) that welding of particles and agglutination would be minimized.

Stage 5. In this stage, gas expansion creates an upward fluidization wave and accelerates chilled pyroclasts. The gas expansion in the dyke caused by the gas venting accelerates the gas into the shattered country rock and produces an upward wave of fluidization that is the major cause of the formation of the diatreme structure (Fig. 2g). The combination of foam disintegration in the upper part of the dyke below the continuous gas phase, and the further foam formation and

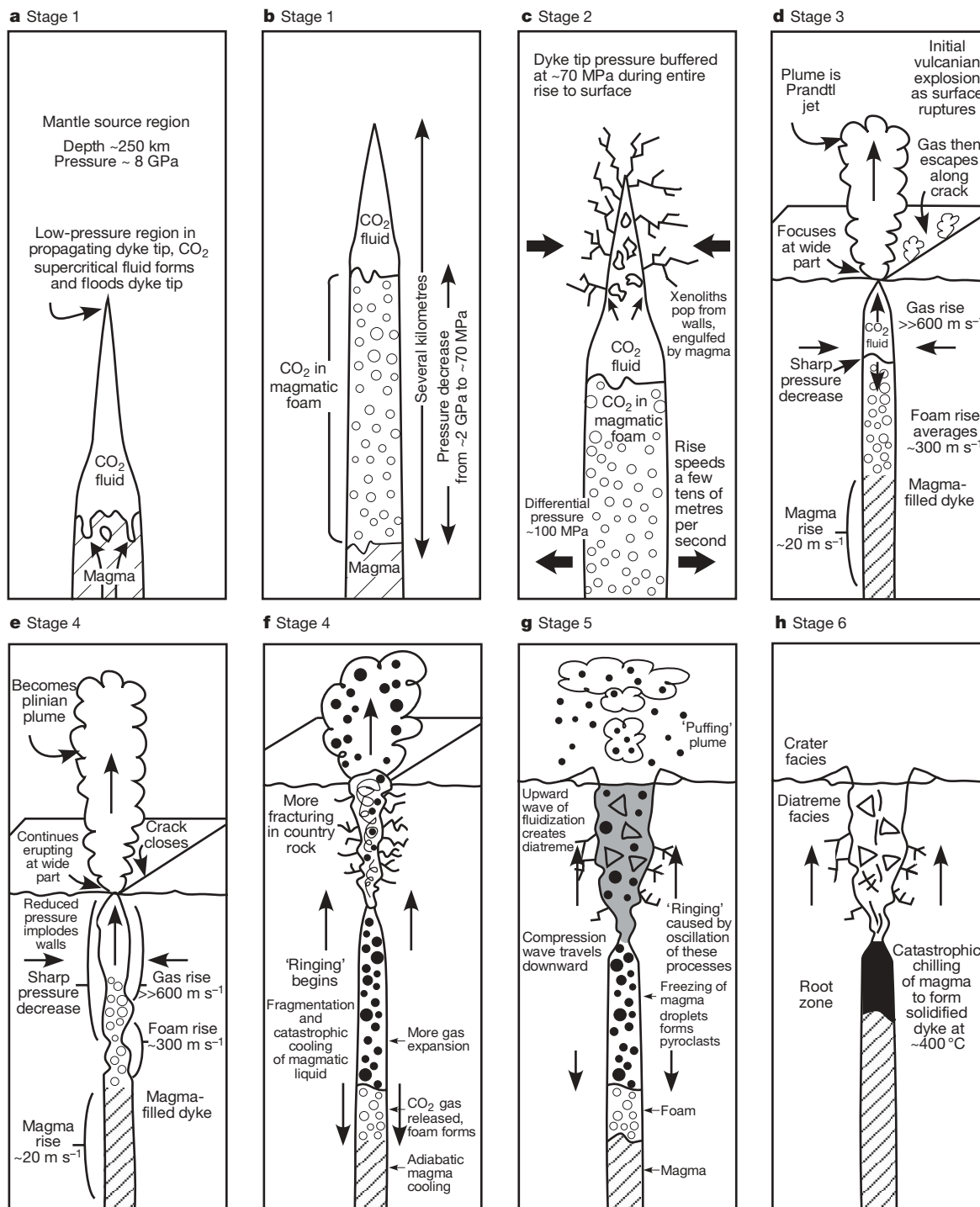


Figure 2 | Sequence of events in the generation, ascent and eruption of kimberlitic magmas and diatreme formation (see text for details).

a, b, Stage 1 involves dyke tip propagation out of a deep source region and CO₂ fluid segregation (**a**). As the dyke tip propagates, disequilibrium produces CO₂/magmatic foam (**b**). **c,** Stage 2 involves dyke ascent and wall fracturing. **d,** In stage 3, the dyke tip breaks the surface, vents CO₂ gas and implodes the walls. **e, f,** In stage 4, gas venting causes a depressurization wave

to propagate into the magmatic foam (**e**) and the underlying magma and country rock (**f**); the diatreme begins to form. **g,** In stage 5, gas expansion creates an upward fluidization wave and accelerates chilled pyroclasts. The vent clogging reduces the pyroclastic escape route, increasing the pressure, which causes 'ringing'. **h,** Stage 6 is the aftermath of the event. The dyke solidifies and the deflated, completed diatreme will be porous and cone-shaped, surrounded by a crater rim of breccia and pyroclastic deposits.

disintegration in the magma below the initial magmatic foam zone, is responsible for the upward fluidization wave. This produces a cold stream of gas through the upper part of the zone, which contains cooled spherules and fine-grained magmatic particles that migrate through the upper fractured zone and vent to the surface (Fig. 2f). Very quickly, however, partial clogging of the pathways and increasing path length cause a growing pressure gradient to be developed in the upper part of the system, and the information that this is occurring propagates downward as a wave of compression (Fig. 2g).

Variations in pressure will cause instabilities in the gas exsolution process and gas flow speed and this will introduce cyclic waves of gas release, pressure changes and venting. A series of fluidization waves (a 'ringing') will propagate through the system. Given the low speed of sound in the complex gas/clast/bubbly liquid mixture, this may last for several tens of minutes. During this time, continual readjustments in the diatreme will be taking place as compression and decompression waves propagate back and forth, and a flow of cool gas transporting particles will permeate the diatreme zone and vent to the surface. This fluidization will cause sorting in the brecciated diatreme zone and will allow settling of large blocks of the country rock from the upper part of the column down into deeper parts of the diatreme. The disruption phase will also serve to modify, distort and destroy evidence of earlier stages of dyke emplacement. During this time, the magma deeper in the dyke itself will quickly undergo catastrophic adiabatic chilling (Fig. 2h) and thus will cease to rise into the diatreme. The deposits on the surface should be characterized by a basal coarse breccia from the initial venting, followed by coarse fragments, xenoliths and lapilli from the initial magmatic foam phase, followed by deposits dominated by products from the second magmatic foam phase (chilled lapilli and ash). The extreme cooling of the magma will inhibit its rise into the vent and any subsequent eruption of surface flows. Diamonds are emplaced into the diatreme facies as part of the kimberlite magma from below the gas and foam zones, where the melt stays at high pressure.

Stage 6. In the aftermath of the event, the deposits will be characterized by a porous, cone-shaped diatreme surrounded by a crater rim of breccia and pyroclastic deposits (Fig. 1). If the diatreme forms in an active groundwater area, a crater lake is likely to form and groundwater will permeate the diatreme, quickly altering the primary mineralogy. Although the predicted surface deposits are similar in some ways to those of tuff cones and maars formed by hydrovolcanic processes, no part of this model requires interaction of the rising dyke with groundwater. Such an interaction could happen, but the very rapid chilling of the magmatic foams minimizes the likelihood of prolonged and repetitive hydrovolcanic eruptions occurring during the formation of diatremes.

Conclusions

This model for the ascent and eruption of CO₂-rich kimberlitic magma (Fig. 2) accounts for the major observational characteristics of kimberlites associated with diatremes^{1,3}, summarized in Fig. 1 and Table 1. The termination of the eruption immediately after diatreme formation, probably within at most a few tens of minutes of the onset of eruption, is a direct consequence of the extreme cooling of magma during the large pressure reductions that occur on venting to the atmosphere. The subsequent very rapid pressure and temperature fluctuations lead to the formation of a diverse suite of rock types in the intrusive deposits that characterize these eruptions. Only limited amounts of pyroclastic materials are expected to be erupted onto the surface. Cases in which abundant CO₂ was lacking in deep melt sources and water was the dominant mantle volatile could result in kimberlite dyke intrusions and eruptions without diatreme formation. Some such events could be of a more protracted nature, building up extensive surface deposits similar to those of traditional basaltic pyroclastic eruptions^{1,15}.

Received 1 November 2006; accepted 19 February 2007.

1. Vespermann, D. & Schmincke, H. in *Encyclopedia of Volcanoes* (ed. Sigurdsson, H.) 683–694 (Academic Press, San Diego, California, 2000).

2. Field, M. & Scott Smith, B. in *Proc. 7th Int. Kimberlite Conf.* (eds Gurney, J. J., Gurney, J. L., Pascoe, M. D. & Richardson, S. H.) Vol. 1, 214–237 (Red Roof Design, Cape Town, 1999).
3. Mitchell, R. H. *Kimberlites* (Plenum, New York, 1986).
4. Dawson, J. Advances in kimberlite geology. *Earth Sci. Rev.* **7**, 187–214 (1971).
5. Hawthorne, J. Model of a kimberlite pipe. *Phys. Chem. Earth* **9**, 1–15 (1975).
6. Clement, C. R. & Reid, A. M. in *Fourth Int. Conf. on Kimberlites and Related Rocks* Vol. 14, 632–646 (Geological Society of Australia, Perth, 1989).
7. Mannard, G. The surface expression of kimberlite pipes. *Proc. Geol. Assoc. Canada* **19**, 15–21 (1968).
8. Wohletz, K. & Sheridan, M. Hydrovolcanic explosions: II, Evolution of basalts tuff rings and tuff cones. *Am. J. Sci.* **283**, 385–413 (1983).
9. Wagner, P. A. *The Diamond Fields of Southern Africa* 1–355 (The Transvaal Leader, Johannesburg, 1914).
10. Kostrovitsky, S. I. *Physical Conditions, Hydraulics and Kinematics of Emplacement of Kimberlite Pipes* 1–95 (Nauka, Novosibirsk, 1976).
11. Reynolds, D. L. Fluidization as a geological process and its bearing on the problem of intrusive granites. *Am. J. Sci.* **255**, 577–613 (1954).
12. McGetchin, T. R. & Ulrich, G. W. Xenoliths in maars and diatremes with inferences for the Moon, Mars, and Venus. *J. Geophys. Res.* **78**, 1833–1853 (1973).
13. McGetchin, T. R., Nikhanj, Y. S. & Chodos, A. A. Carbonatite-kimberlite relations in the Cane Valley diatreme, San Juan County, Utah. *J. Geophys. Res.* **78**, 1854–1869 (1973).
14. Wyllie, P. J. The origin of kimberlite. *J. Geophys. Res.* **85**, 6902–6910 (1980).
15. Skinner, E. & Marsh, J. Distinct kimberlite pipe classes with contrasting eruption processes. *Lithos* **76**, 183–200 (2004).
16. Lorenz, V. in *Physics and Chemistry of the Earth* Vol. 9, 17–27 (Pergamon Press, Oxford, 1975).
17. Lorenz, V., Zimanowski, B., Buttner, R. & Kurszlaukis, S. in *Proc. VII Int. Kimberlite Conf.* (eds Gurney, J. J., Gurney, J. L., Pascoe, M. D. & Richardson, S. H.) Vol. 2, 522–528 (Red Roof Design, Cape Town, 1999).
18. Wolfe, J. A. Fluidization versus phreatomagmatic explosions in breccia pipes. *Econ. Geol.* **75**, 1105–1111 (1980).
19. Clement, C. R. The origin and infilling of kimberlite pipes. In *Extended Abstracts, Kimberlite Symposium II*, Cambridge (Geology Dept, DeBeers Consolidated Mines Ltd, 1979).
20. Clement, C. R. *A Comparative Geological Study of Some Major Kimberlite Pipes in the Northern Cape and Orange Free State*. PhD thesis, Univ. Cape Town (1982).
21. Green, H. W. & Gueguen, Y. Origin of kimberlite pipes by diapiric upwelling in the upper mantle. *Nature* **249**, 617–620 (1974).
22. Kramers, J. D., Smith, C. B., Lock, N. P., Harmon, R. S., Boyd, F. R. Can kimberlites be generated from an ordinary mantle? *Nature* **291**, 53–56 (1981).
23. Wyllie, P. in *Proc. 2nd Int. Kimberlite Conf.* (eds Meyer, H. and Boyd, F. R.) Vol. 1, 319–329 (AGU, Washington DC, 1979).
24. Wyllie, P. & Ryabchikov, I. D. Volatile components, magmas, and critical fluids in upwelling mantle. *J. Petrol.* **41**, 1195–1206 (2000).
25. Anderson, O. L. in *Proc. 2nd Int. Kimberlite Conf.* (eds Meyer, H. and Boyd, F. R.) Vol. 1, 344–353 (AGU, Washington DC, 1979).
26. Sleep, N. H. Tapping of melt by veins and dikes. *J. Geophys. Res.* **93**, 10255–10272 (1988).
27. Lister, J. R. Buoyancy-driven fluid fracture: The effects of material toughness and of low-viscosity precursors. *J. Fluid Mech.* **210**, 263–280 (1990).
28. Rubin, A. M. Dikes vs. diapirs in viscoelastic rock. *Earth Planet. Sci. Lett.* **119**, 641–659 (1993).
29. Wyllie, P. J. & Huang, W. L. Influence of mantle CO₂ in the generation of carbonatites and kimberlites. *Nature* **257**, 297–299 (1975).
30. Wilson, L. & Head, J. W. Deep generation of magmatic gas on the Moon and implications for pyroclastic eruptions. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL016082 (2003).
31. Wilson, L., Sparks, R. S. J. & Walker, G. P. L. Explosive volcanic eruptions—IV. The control of magma chamber and conduit geometry on eruption column behavior. *Geophys. J. R. Astron. Soc.* **63**, 117–148 (1980).
32. Lister, J. R. Buoyancy-driven fluid fracture: Similarity solutions for the horizontal and vertical propagation of fluid-filled cracks. *J. Fluid Mech.* **217**, 213–239 (1990).
33. Hess, P. *Origins of Igneous Rocks* 1–336 (Harvard Univ. Press, Cambridge, Massachusetts, 1989).
34. Kieffer, S. W. Blast dynamics at Mount St. Helens on 18 May 1980. *Nature* **291**, 568–570 (1981).
35. Wilson, L. & Head, J. W. Ascent and eruption of basaltic magma on the Earth and Moon. *J. Geophys. Res.* **86**, 2971–3001 (1981).
36. Canil, D. & Fedortchouk, Y. Garnet dissolution and the emplacement of kimberlites. *Earth Planet. Sci. Lett.* **167**, 227–237 (1999).
37. Kieffer, S. W. Sound speed in liquid-gas mixtures: Water-air and water-steam. *J. Geophys. Res.* **82**, 2895–2904 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by a grant from the National Aeronautics and Space Administration to J.W.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.W.H. (james_head@brown.edu).

ARTICLES

The structure of a plant photosystem I supercomplex at 3.4 Å resolution

Alexey Amunts¹, Omri Drory¹ & Nathan Nelson¹

All higher organisms on Earth receive energy directly or indirectly from oxygenic photosynthesis performed by plants, green algae and cyanobacteria. Photosystem I (PSI) is a supercomplex of a reaction centre and light-harvesting complexes. It generates the most negative redox potential in nature, and thus largely determines the global amount of enthalpy in living systems. We report the structure of plant PSI at 3.4 Å resolution, revealing 17 protein subunits. PsaN was identified in the luminal side of the supercomplex, and most of the amino acids in the reaction centre were traced. The crystal structure of PSI provides a picture at near atomic detail of 11 out of 12 protein subunits of the reaction centre. At this level, 168 chlorophylls (65 assigned with orientations for Q_x and Q_y transition dipole moments), 2 phylloquinones, 3 Fe_4S_4 clusters and 5 carotenoids are described. This structural information extends the understanding of the most efficient nano-photochemical machine in nature.

Oxygenic photosynthesis is the principal producer of both oxygen and organic matter on Earth^{1–3}. Water, the electron donor for this process, is oxidized to O_2 and four protons by PSII. The electrons that have been extracted from water are shuttled through a quinone pool and the cytochrome b_6f complex to plastocyanin—a small, soluble, copper-containing protein. Solar energy that has been absorbed by PSI induces the translocation of an electron from plastocyanin at the inner face of the membrane (thylakoid lumen) to ferredoxin on the opposite side (stroma). PSI generates the most negative redox potential in nature (-1 V), and thus largely determines the global amount of enthalpy in living systems. The structures of three of the four complexes that catalyse oxygenic photosynthesis in cyanobacteria have been solved at relatively high resolution, and the position of most of their amino acids and prosthetic groups has been defined^{4–8}. Thus, the architecture of oxygenic photosynthesis in cyanobacteria has largely been determined. The structure of cytochrome b_6f complex from chloroplasts of the algae *Chlamydomonas reinhardtii* has also been solved at high resolution, and has remarkable similarity to the cyanobacterial complex⁹. Recently, two high-resolution structures of light-harvesting complexes of PSII from higher plants were published^{10,11}.

In a previous study, we crystallized and determined the structure of plant PSI at an intermediate resolution of 4.4 Å^{12,13}. A model of α -carbon chains of 16 subunits, 45 transmembrane helices, 2 phylloquinones, 3 iron–sulphur clusters and 167 chlorophyll molecules was presented. To understand better the role of the protein subunits and individual amino acids in the binding and organization of the various cofactors, as well as the interaction between the cofactors, we have determined the X-ray crystallographic structure of PSI from pea (*Pisum sativum*) at 3.4 Å resolution, and we describe a near atomic model of the system.

Features revealed by the new structure

The new crystal form contains one complex in the asymmetric unit, with unit cell parameters $a = 124.87$ Å, $b = 187.27$ Å, $c = 131.96$ Å, $\beta = 91.03^\circ$ (Supplementary Information and Supplementary Fig. 1). In the electron density map, all the previously detected 16 subunits of PSI were identified, and PsaN was also detected (Fig. 1). Twelve of the

subunits (PsaA, B, C, D, E, F, G, H, I, J, L, and N; ref. 14) were interpreted with the known amino acid sequences from pea and *Arabidopsis thaliana* plants (Supplementary Tables 1 and 2). The position of 3,038 out of 3,443 predicted amino acids was assigned; for 2,909 of them, side chains were built into the model. Part of PsaK could be modelled as polyalanine, because the electron density was not as well defined in this part of the supercomplex. Subunit O (ref. 15), which was identified in preparations of plant PSI, was not

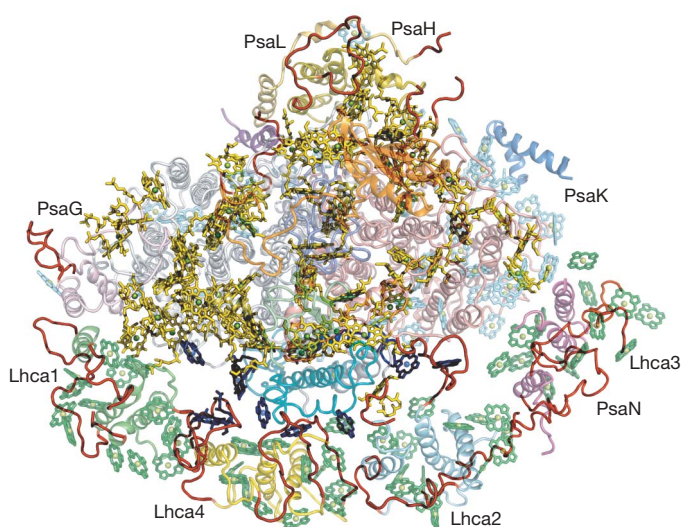


Figure 1 | The structural model of plant photosystem I at 3.4 Å resolution.

View from the stroma of the structure of plant PSI. Novel structural elements that are not present in the previous model are shown as red ribbon structures. Chlorophylls with detected phytol side chains, revealing the orientation of the Q_x and Q_y transition dipole moments, are yellow. The rest of the reaction centre chlorophylls are cyan, gap chlorophylls are blue and chlorophylls of LHCI are green. The positions of PsaG, H, K, L and N, as well as the various LHCI monomers, are indicated. Each individual subunit is coloured differently.

¹Department of Biochemistry, The George S. Wise Faculty of Life Sciences, The Daniella Rich Institute for Structural Biology, Tel Aviv University, Tel Aviv, 69978, Israel.

present in the crystal and there was no space for it to be accommodated. Subunit P (ref. 16) could not be positively identified, but unassigned electron density close to PsaH could possibly be assigned to this subunit. PsaN, which was not detected in the 4.4 Å structure, was identified and its amino and carboxy termini were determined by partial amino acid tracing (Fig. 1). The entire length of PsaG was traced and the interaction of its loop with PsaB gives credence to the conclusions of a recent mutational analysis¹⁷. The positive amino acids that are present in the loop are protected from proteases by close proximity to the membrane (Lys 59), salt bridge formation (Arg 50 with Glu 306 of PsaB) and tight interaction with PsaB (Lys 52). A major part of Lhca1–4 was assigned using the corresponding amino acids, but parts of the loops that could be detected were modelled by amino acids without side chains. Tracing amino acids of the transmembrane helices of Lhca1–4 allowed positive identification of the four light-harvesting subunits. For the 168 chlorophylls, the position and Q_z orientation of the head groups were determined. For 65 of the chlorophylls, the ring substituents (and part of their phytol side chains) could be modelled into the electron density map, revealing the orientation of the Q_x and Q_y transition dipole moments.

The PSI reaction centre complex

The two principal subunits of the reaction centre, PsaA and PsaB, share similarities in their amino acid sequences and constitute a pseudosymmetric structure that evolved from an ancient homodimeric assembly^{3,18}. Together, the subunits harbour the electron transport chain (ETC), which is the heart of PSI and functions in the photoelectrochemical reaction of the system. In addition, two reaction centre proteins are exclusively present in plants and green algae (subunits G and H). The position and shape of PsaH conform well to its proposed role as a docking site for light harvesting complex (LHC)II (ref. 19). Part of the N terminus that was not detected in the previous report was traced in the current study (Fig. 1) and was found to form an additional surface that may be used for controlled binding of LHCII and other auxiliary factors. On the opposite side of the reaction centre, PsaG and its two tilting transmembrane helices contribute most of the contact surface area for association with LHCI (ref. 13). Twenty amino acids of the PsaG C terminus (Fig. 1) protrude out of the generally compact structure in this area; this may provide a binding surface for other membrane complexes such as the

cytochrome *b₆f* complex. The electron densities at the centre were clear enough to correct potential sequencing mistakes (Supplementary Fig. 2). For example, there is an arginine residue in the sequence of pea chloroplasts at position 220 of PsaA; in all other plants, there is a glycine residue at the same position. The electron density obtained here leaves little doubt that, in PsaA from pea, there is also a glycine residue at this position. On the luminal side, the most noticeable distinction between plant and cyanobacterial reaction centres is the helix–loop–helix motif contributed by the longer N-terminal domain of plant PsaF (Fig. 2a and Supplementary Fig. 3). This domain enables more efficient plastocyanin binding in plants and, as a result, two orders of magnitude faster electron transfer from the copper protein to P700 (ref. 13). On the stromal side of PSI, where ferredoxin and ferredoxin-NADP-reductase bind, almost all amino acids of PsaC, PsaD and PsaE were traced. The hypothetical amino acid chain T was not visible in the current electron densities.

The PsaA/B part of the ETC is formed by six chlorophylls, two phylloquinones and one out of the three Fe₄S₄ clusters of PSI (Fig. 2a). The chlorophylls and quinones are arranged in two branches (A and B) as pairs of molecules related by a pseudo-C₂ axis. The two chlorophylls that constitute the P700, and most other chlorophylls at the centre, seem to be identically positioned to those reported for the cyanobacterium *Synechococcus elongatus*⁴. The amino acids involved in Mg²⁺ coordination and hydrogen bonding to the second and third Chl a pairs of the ETC are strictly conserved between the PsaA and PsaB, in species from cyanobacteria to higher plants⁴. One of the unsolved questions in the mechanism of the ETC is whether the Chl a from one, or both (PsaA and PsaB), branches are active in electron transport⁸. Several electron paramagnetic resonance (EPR) experiments with cyanobacterial PSI suggest that most of the electron transfer is conducted by the A branch²⁰; however, fast spectroscopic experiments with algae suggest an equal share of the two branches in electron transfer²¹. Even though small changes in the cofactor positions were detected, the current resolution does not permit a definite answer to this question.

Chlorophylls and carotenes at the reaction centre

The core complex (reaction centre) of plant PSI contains approximately 100 chlorophyll molecules, of which a vast majority maintain an almost identical position, as in cyanobacterial PSI¹³. For 65

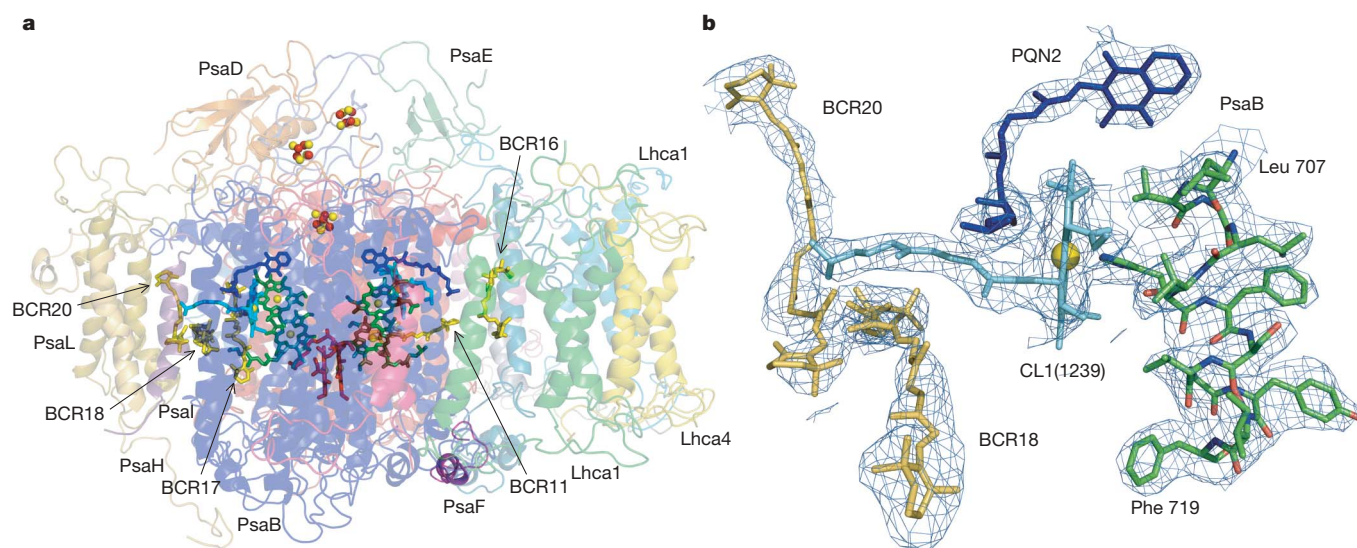


Figure 2 | Position of β -carotenes in relation to the ETC. **a**, The ETC, two additional chlorophylls and three β -carotene molecules are depicted on the background of the subunit structure of PSI. Red, P700 chlorophylls 9010 and 9011; green, ETC chlorophylls 9012, 9013, 9022 and 9023; cyan, chlorophylls 1239 and 1140; yellow, β -carotenes 6011 and 6018; and blue, phylloquinones.

The three sulphur–iron clusters are shown as spheres in red (iron) and yellow (sulphur). **b**, The 2F_o-F_c electron density map (1 σ), covering β -carotenes 6018 and 6020 (BCR18 and BCR20), chlorophyll 1239 (CL1(1239)), and phylloquinone 5002 (PQN2), as well as part of PsaB (amino acids 707–719). Colour codes correspond to Fig. 3a.

molecules, the electron densities were good enough to trace part of their phytol side chains, revealing the Q_x and Q_y transition dipole moments (Fig. 1). These chlorophylls exhibit a remarkable conservation in their position and orientation compared to those of *S. elongatus*⁴. As expected, the chlorophylls that are coordinated to subunits M and X in cyanobacteria were missing in the plant reaction centre.

The construction of our model was aided by a theoretical atomic model of plant PSI that was built by combining the low-resolution model of plant PSI¹³, the high resolution structure of the cyanobacteria PSI⁴, and a new approach of molecular dynamics²². The position and orientation of most chlorophyll molecules of the current model are in accordance with the theoretical model. However, the refined position of several chlorophyll molecules was significantly different from the theoretical model. Plant PSI contains 19 chlorophyll molecules, including 9 gap chlorophylls (Fig. 1), that are not present in the cyanobacterial reaction centre and are not part of the LHCI monomers^{23,24}. In the neighbourhood of PsaN, we modelled four chlorophyll molecules, some of which may have an important role in excitation energy transfer from LHCII to PSI (see below). This side of PSI is poorly resolved not only in the larger plant PSI but also in the high-resolution PSI of *S. elongatus*⁴.

Over 20 carotene molecules are expected to be present in plant PSI (ref. 22). Relatively good electron densities allowed for the assignment of five β -carotene (BCR) molecules in various locations of the reaction centre (Fig. 2). BCR11 is situated in a strategic location in the vicinity of the proposed excitation energy transfer pathway from the reaction centre antenna to the ETC—approximately 10 Å away from P700—and one of its poles is 3 Å away from Chl1126 and thus adjacent to Trp 747/A and His 389/A. The second pole is 6 Å away from Chl1230, which is coordinated by His 439/B, and 4.5 Å from Chl1229, and is in the vicinity of Phe 90/F. This BCR is located at a similar position to β -carotene in the cyanobacterial PSI and in the proposed theoretical model of plant PSI²². However, BCR16 moved considerably from its position in *S. elongatus*⁴—because subunit X is not present in plant PSI, the chlorophyll molecule that it coordinated is missing and two gap chlorophyll molecules (Chl1302 and Chl1303) were added to the plant complex. One of the poles of BCR16 is as close as 3.3 Å to Chl1303, and the other pole is situated 5 and 4.3 Å from Trp 99/F and Trp 136/F, respectively. Most of the β -carotenes in cyanobacteria were located in pockets of hydrophobic residues that are highly conserved between cyanobacteria and plants. This arrangement may protect not only from triplets that are formed by the

pigment molecules but also from destruction of aromatic amino acids by ultraviolet light. BCR17 is coordinated by Trp 648/B and Trp 646/B; Phe 652/B and Phe 719/B are situated in its vicinity (Fig. 2). BCR17, 18 and 20 are close to each other, suggesting potential radiation damage in this part of the reaction centre.

Light-harvesting complex I

The LHCI belt, with its associated chlorophylls, is the most prominent addition to PSI structure by plants and green algae. The LHCI belt contributes a mass of 160 kDa out of approximately 600 kDa in PSI. LHCI is composed of four nuclear gene products (Lhca1–Lhca4) that are 20–24 kDa polypeptides and belong to the LHC family of chlorophyll a/b binding proteins. The archetype of this family, and the most abundant membrane protein in nature, is the major LHCII protein (Lhcb1–2), the structure of which was recently elucidated by X-ray crystallography at 2.7 (ref. 10) and 2.5 Å (ref. 11) resolution. The identification of the four Lhca proteins in plant PSI was based on a large body of biochemical experiments^{13,25–27}. We were able to positively identify the four LHCI units as Lhca1, Lhca4, Lhca2 and Lhca3, starting at the G-pole of the reaction centre, respectively. This was achieved by assigning electron densities to corresponding amino acids with large side chains that are unique to each LHCI protein. Binding of LHCI to the reaction centre is asymmetric, namely, much stronger on the G-pole than on the K-pole of the core (Fig. 1). The other LHCI proteins interact with the core mainly through small binding surfaces at their stromally exposed regions (Fig. 1). Lhca4 binds to Psaf, Lhca2 associates weakly with Psaj, and Lhca3 binds weakly to Psaa.

Even though the light harvesting chlorophyll a/b binding proteins that constitute the peripheral antennas of PSI (LHCI) and PSII (LHCII) share sequence and structural homology, their oligomeric states vary considerably. Whereas LHCI proteins assemble into dimers, the light harvesting proteins that associate with PSII form either trimers (LHCII), or monomers as minor antenna members CP24, CP26 and CP29 (ref. 1). The current model permitted a closer look at the dimer formation and mode of interaction between the LHCI monomers and reaction centre subunits (Fig. 3a). Both Lhca1–4 and LHCII bind 14 chlorophyll a/b molecules each (Fig. 1), and possess the LHCII general fold^{10,11,13,28}. The absorption peak of the 'bulk' chlorophylls of LHCI proteins is also shifted to lower energies compared with LHCII^{13,24,29}. Dimerization in LHCI is mediated by relatively small contact surfaces at the luminal side by the C terminus and at the stromal side of the N-terminal domain of the Lhca

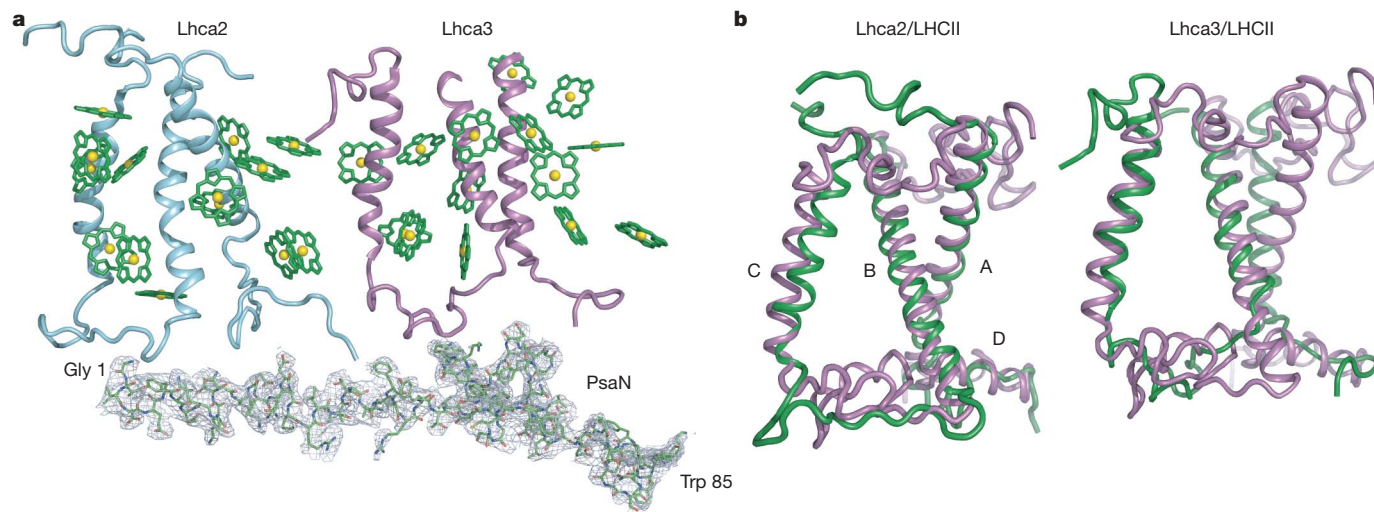


Figure 3 | The position of PsaN in relation to Lhca2 and Lhca3, and the unique fold of Lhca3. **a**, The $2F_oF_c$ (1σ) electron density map covering PsaN, and the structure of the Lhca2–Lhca3 heterodimer. Cyan, Lhca2; magenta,

Lhca3; green, chlorophylls; yellow, magnesium atoms. **b**, Left panel, superposition of LHCII (magenta) on Lhca2 (green); right panel, superposition of LHCII (magenta) on Lhca3 (green).

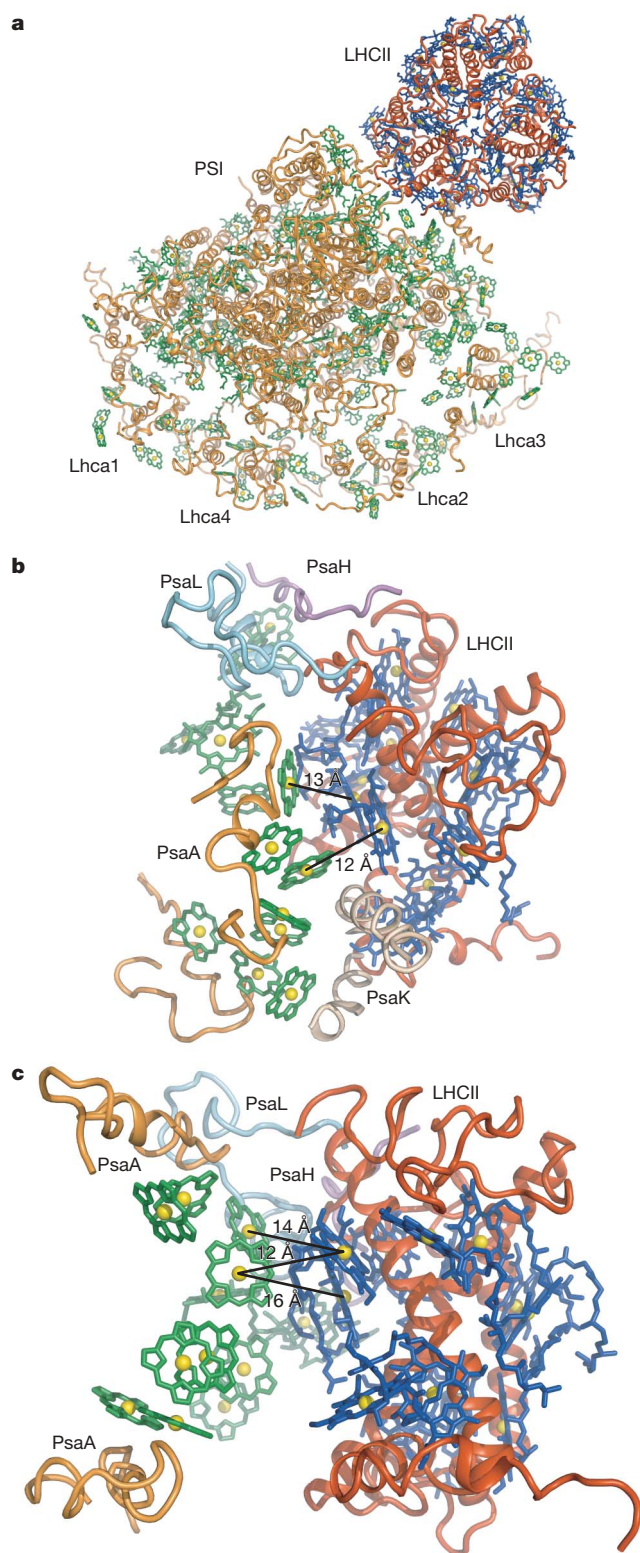


Figure 4 | Model for PSI–LHCII interactions. The structural models of plant PSI and the LHCII trimer were fitted by accommodating a possible binding site at the PsaK side. In this position, LHCII could be readily crosslinked to subunits PsaL and PsaH, but it is too far to crosslink with PsaI. The initial docking was made by modified PatchDock software (<http://bioinfo3d.cs.tau.ac.il>). The fit was manually improved to better agree with experimental data. **a**, A view from the stroma of plant PSI together with a LHCII trimer (Protein Data Bank code, 2BHW; ref. 11). The PSI complex (orange), PSI chlorophylls (green), LHCII trimer (red), LHCII chlorophylls (blue) and chlorophyll magnesium atoms (yellow) are shown. **b**, **c**, An enlarged view from the stroma (**b**) and a view along the membrane plane

proteins¹³ (Fig. 3a). A similar mode of association is observed between LHCI dimers, which allows all LHCI proteins to have their wider side turned to the reaction centre, enabling the maximum number of chlorophylls to face the core²⁴. This arrangement results in relatively long distances between the membrane domains of adjacent monomers.

The half-moon shape of LHCI and the relatively loose and flexible coupling among its monomers may serve two important functions: (1) achieving the most efficient light harvesting and excitation energy migration, and (2) providing the basis for coping with ever-changing light intensities. Increased light intensities result in a sharp decrease in antenna size associated with the vulnerable PSII³⁰; however, such an effect was not observed in LHCI. The composition—rather than the size—of this peripheral antenna varies with intensity³¹. It is possible that replacement of the Lhca2–Lhca3 heterodimer with an Lhca3–Lhca3 dimer results in longer trapping times, decreased efficiency in energy migration to the reaction centre, and dissipation of energy localized on Lhca3 by carotenoids.

The conservation of amino acid sequences between LHCII and the four LHCI monomers was found to be relatively poor. The four LHCI monomers share the LHCII general fold; that is, two long, tilted, intertwined transmembrane helices (A and B) and a shorter one roughly perpendicular to the membrane (C). The conservation of the fourth helix (D) was not apparent, and the fine structure of Lhca3 was found to be significantly different from the other three LHCI monomers (Fig. 3b). In Lhca3, helices A and B were much closer to each other, were almost parallel and were less intertwined. Consequently, Lhca3 bound its chlorophylls in a somewhat different position than the other three LHCI monomers. Recent reconstitution experiments clearly demonstrate that Lhca3 shares a similar fold to the others and binds its chlorophylls in the same fashion³². It therefore seems that when Lhca3 assembles by itself, it results in a similar structure to the others, but when it assembles in the context of the rest of PSI, it adopts a different conformation.

Subunit N

An electron density was identified at the luminal side of PSI close to Lhca2 and Lhca3. Because the volume of the density and the position at the lumen could be ascribed only to PsaN, we modelled into it the amino acids of this subunit (Fig. 3a). We used the available sequence from the plant *Phaseolus vulgaris* (accession number AA049652), which is expected to be almost identical to that from the related pea plants and differs from that of *A. thaliana* by only nine conservative substitutions. The electron densities in the region of PsaN are well defined, but at 3.4 Å resolution did not allow tracing of the amino acids sequentially. The positions of N and C termini were determined by secondary structural analysis, which suggested an approximately 30-amino-acid-long, predominantly α -helical stretch, at the N-terminal part of PsaN. In line with biochemical evidence^{33,34}, the structure of PsaN exhibits weak interactions with Lhca2 and Lhca3 (Fig. 3a).

Subunit N was first identified in a high ionic strength wash of PSI preparation from spinach chloroplasts^{33,34}. The complete sequence of PsaN was first reported from barley³⁵ and has subsequently been reported in several other species of higher plants and green algae. An extensive crosslinking study revealed minimal interaction between PsaN and other small PSI subunits³⁶. Putative crosslinking products between PsaN and PsaG and between PsaN and PsaF have been found, and specific inactivation of a nuclear gene encoding a PSI subunit N in *A. thaliana* plants has been reported³⁷. The lack of PsaN

(**b**), of the suggested PSI–LHCII interaction site. LHCII (red) interacts with PSI subunits PsaH (magenta), PsaL (cyan), PsaA (orange) and PsaK (brown). Two LHCII α -chlorophylls (numbers 602 and 607, in blue) have energy transfer distances of 13 Å to 12 Å from two PSI chlorophylls (numbers 1151 and 1153, in green), respectively, that are coordinated by PsaA.

results in effects on plant growth and development under suboptimal conditions. Under standard growth conditions, plants compensate for deficient PSI by increasing the relative PSI content. It was proposed that PsaN is necessary for efficient interaction of PSI with plastocyanin. Our structure reveals no direct interaction with PsaG, PsaF or plastocyanin, but, as frequently occurred, indirect effects of the missing subunit might lead to the reported effects. Recent findings indicate that Lhca5 assembles onto Lhca2 (ref. 38); the position of PsaN (Fig. 4a) suggests it is involved in this process.

The PSI–LHCII supercomplex

The determined structure of plant photosystem I (PSI) provides the first relatively high-resolution structural model of a supercomplex containing a reaction centre and its peripheral antenna. This highly efficient nano-photoelectric machine is expected to interact with other proteins in a regulated and efficient manner. The most important interaction of PSI at the membrane level is with LHCII during state transition⁸. Plants adapt to changes in light quality by redistributing excitation energy between the two photosystems to enhance photosynthetic yield³⁹. At high light intensities, LHCII migrates from PSII to PSI⁴⁰. State transitions in higher plants are limited. In state II, additional light harvesting by PSI does not exceed 20% (ref. 41) and corresponds to an addition of up to a single LHCII trimer to PSI, forming a supercomplex of reaction-centre–LHCI–LHCII. Numerous experiments conducted in higher plants, including single particle analysis^{42,43}, support this model, but the detailed mechanism of state transitions remains unclear^{44–46}. Even though LHCII is likely to interact with PSI at the PsaK side (which is less well resolved than the PsaG side), we superimposed LHCII on the current structure and found only one position with good fit (Fig. 4). In this model, only one of the three LHCII monomers was found to interact with the reaction centre, indicating that monomers may also fit well to the binding site. Recently it was reported that, in *C. reinhardtii*, CP29 associates with PsaH of PSI and was proposed to act as a docking site for LHCII during state transition⁴⁵. Our model for the complex PSI–LHCII is compatible with the possibility that a similar interaction may take place in higher plants.

The complexity of PSI belies its efficiency: almost every photon absorbed by the PSI complex is used to drive electron transport. It is remarkable that PSI exhibits a quantum yield of nearly 1 (refs 47, 48), and every captured photon is eventually trapped and results in electron translocation. The structural information on the proteins, the cofactors and their interactions that is described in this work provides a step towards understanding how the unprecedented high quantum-yield of PSI in light capturing and electron transfer is achieved.

METHODS

PSI was isolated from pea and crystallized by a modified procedure that was described previously^{12,13,49}. Changes to the procedure included additional sucrose gradient centrifugation in a SW60 rotor (Beckman) at 57,000 r.p.m. for 4 h, as well as substitution of citrate by succinate and adjustment of the pH to 6.0 for crystallization (see Supplementary Information). A crystal form was obtained by prolonged incubation with cryo solution containing a high concentration of PEG 6000. A similar procedure was used previously with crystals containing high water content (see Supplementary Information). The cryo solution contained 22 mM citrate, 22.5 mM MES/bis Tris (pH 6.7), 0.5% PEG 400, and 40% PEG6000. The crystals were washed in the same solution containing only 20% PEG 6000 and after 1 h were placed in the above solution for approximately 1 week. Before mounting, the crystals were incubated at room temperature for 1 day and frozen by liquid nitrogen or a nitrogen stream at 100 K. Detailed information on the X-ray data collection, evaluation and refinement at 3.4 Å resolution are provided in the Supplementary Information. As shown in Supplementary Fig. 1, the crystal lattice was changed from two PSI complexes in the asymmetric unit to a single complex maintaining the symmetry of P21.

Received 1 December 2006; accepted 19 February 2007.

- Barber, J. Engine of life and big bang of evolution: a personal perspective. *Photosynth. Res.* **80**, 137–155 (2004).

- Nelson, N. & Ben-Shem, A. The complex architecture of oxygenic photosynthesis. *Nature Rev. Mol. Cell Biol.* **5**, 971–982 (2004).
- Nelson, N. & Ben-Shem, A. The structure of photosystem I and evolution of photosynthesis. *Bioessays* **27**, 914–922 (2005).
- Jordan, P. et al. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).
- Kurusu, G., Zhang, H., Smith, J. L. & Cramer, W. A. Structure of the cytochrome b₆f complex of oxygenic photosynthesis: tuning the cavity. *Science* **302**, 1009–1014 (2003).
- Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. Architecture of the photosynthetic oxygen-evolving center. *Science* **303**, 1831–1838 (2004).
- Loll, B., Kern, J., Saenger, W., Zouni, A. & Biesiadka, J. Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature* **438**, 1040–1044 (2005).
- Nelson, N. & Yocum, C. Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* **57**, 521–565 (2006).
- Stroebel, D., Choquet, Y., Popot, J. L. & Picot, D. An atypical haem in the cytochrome b₆f complex. *Nature* **426**, 413–418 (2003).
- Liu, Z. et al. Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* **428**, 287–292 (2004).
- Standfuss, J., Terwisscha van Scheltinga, A. C., Lamborghini, M. & Kuhlbrandt, W. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 Å resolution. *EMBO J.* **24**, 919–928 (2005).
- Ben-Shem, A., Nelson, N. & Frolov, F. Crystallization and initial X-ray diffraction studies of higher plant photosystem I. *Acta Crystallogr. D* **59**, 1824–1827 (2003).
- Ben-Shem, A., Frolov, F. & Nelson, N. The crystal structure of plant photosystem I. *Nature* **426**, 630–635 (2003).
- Scheller, H. V., Jensen, P. E., Haldrup, A., Lunde, C. & Knoetzel, J. Role of subunits in eukaryotic Photosystem I. *Biochim. Biophys. Acta* **1507**, 41–60 (2001).
- Jensen, P. E., Haldrup, A., Zhang, S. & Scheller, H. V. The PSI-O subunit of plant photosystem I is involved in balancing the excitation pressure between the two photosystems. *J. Biol. Chem.* **279**, 24212–24217 (2004).
- Khrouchtchova, A. et al. A previously found thylakoid membrane protein of 14 kDa (TMP14) is a novel subunit of plant photosystem I and is designated PSI-P. *FEBS Lett.* **579**, 4808–4812 (2005).
- Zygadlo, A., Robinson, C., Scheller, H. V., Mant, A. & Jensen, P. E. The properties of the positively charged loop region in PSI-G are essential for its “spontaneous” insertion into thylakoids and rapid assembly into the photosystem I complex. *J. Biol. Chem.* **281**, 10548–10554 (2006).
- Ben-Shem, A., Frolov, F. & Nelson, N. Evolution of Photosystem I—from symmetry through pseudosymmetry to asymmetry. *FEBS Lett.* **564**, 274–280 (2004).
- Lunde, C. P., Jensen, P. E., Haldrup, A., Knoetzel, J. & Scheller, H. V. The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis. *Nature* **408**, 613–615 (2000).
- Dashdorj, N., Xu, W., Cohen, R. O., Golbeck, J. H. & Savikhin, S. Asymmetric electron transfer in cyanobacterial Photosystem I: charge separation and secondary electron transfer dynamics of mutations near the primary electron acceptor A0. *Biophys. J.* **88**, 1238–1249 (2005).
- Joliot, P. & Joliot, A. *In vivo* analysis of the electron transfer within photosystem I: are the two phyloquinones involved? *Biochemistry* **38**, 11130–11136 (1999).
- Jolley, C., Ben-Shem, A., Nelson, N. & Fromme, P. Structure of plant photosystem I revealed by theoretical modeling. *J. Biol. Chem.* **280**, 33627–33636 (2005).
- Morosinotto, T., Ballottari, M., Klimmek, F., Jansson, S. & Bassi, R. The association of the antenna system to photosystem I in higher plants. Cooperative interactions stabilize the supramolecular complex and enhance red-shifted spectral forms. *J. Biol. Chem.* **280**, 31050–31058 (2005).
- Ben-Shem, A., Frolov, F. & Nelson, N. Light-harvesting features revealed by the structure of plant photosystem I. *Photosynth. Res.* **81**, 239–250 (2004).
- Ganeteg, U., Strand, A., Gustafsson, P. & Jansson, S. The properties of the chlorophyll a/b-binding proteins Lhca2 and Lhca3 studied *in vivo* using antisense inhibition. *Plant Physiol.* **127**, 150–158 (2001).
- Schmid, V. H., Paulsen, H. & Rupprecht, J. Identification of N- and C-terminal amino acids of Lhca1 and Lhca4 required for formation of the heterodimeric peripheral photosystem I antenna LHCI-730. *Biochemistry* **41**, 9126–9131 (2002).
- Ihalainen, J. A. et al. Pigment organization and energy transfer dynamics in isolated photosystem I (PSI) complexes from *Arabidopsis thaliana* depleted of the PSI-G, PSI-K, PSI-L, or PSI-N subunit. *Biophys. J.* **83**, 2190–2201 (2002).
- Durnford, D. G. et al. A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J. Mol. Evol.* **48**, 59–68 (1999).
- Morosinotto, T., Castelletti, S., Breton, J., Bassi, R. & Croce, R. Mutation analysis of Lhca1 antenna complex. Low energy absorption forms originate from pigment–pigment interactions. *J. Biol. Chem.* **277**, 36253–36261 (2002).
- Anderson, J. M., Chow, W. S. & Park, Y.-I. The grand design of photosynthesis: acclimation of the photosynthetic apparatus to environmental cues. *Photosynth. Res.* **46**, 129–139 (1995).
- Bailey, S., Walters, R. G., Jansson, S. & Horton, P. Acclimation of *Arabidopsis thaliana* to the light environment: the existence of separate low light and high light responses. *Planta* **213**, 794–801 (2001).
- Mozzo, M., Morosinotto, T., Bassi, R. & Croce, R. Probing the structure of Lhca3 by mutation analysis. *Biochim. Biophys. Acta* **1757**, 1607–1613 (2006).

33. Ikeuchi, M. & Inoue, Y. Two new components of 9 and 14 kDa from spinach photosystem I complex. *FEBS Lett.* **280**, 332–334 (1991).
34. He, W. Z. & Malkin, R. Specific release of a 9-kDa extrinsic polypeptide of photosystem I from spinach chloroplasts by salt washing. *FEBS Lett.* **308**, 298–300 (1992).
35. Knoetzel, J. & Simpson, D. J. The primary structure of a cDNA for PsaN, encoding an extrinsic luminal polypeptide of barley photosystem I. *Plant Mol. Biol.* **22**, 337–345 (1993).
36. Jansson, S., Andersen, B. & Scheller, H. V. Nearest-neighbor analysis of higher-plant photosystem I holocomplex. *Plant Physiol.* **112**, 409–420 (1996).
37. Haldrup, A., Naver, H. & Scheller, H. V. The interaction between plastocyanin and photosystem I is inefficient in transgenic *Arabidopsis* plants lacking the PSI-N subunit of photosystem I. *Plant J.* **17**, 689–698 (1999).
38. Lucinski, R., Schmid, V.H., Jansson, S. & Klimmek, F. Lhca5 interaction with plant photosystem I. *FEBS Lett.* **580**, 6485–6488 (2006); published online 7 November 2006.
39. Bellafiore, S., Barneche, F., Peltier, G. & Rochaix, J. D. State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. *Nature* **433**, 892–895 (2005).
40. Kyle, D. J., Staehelin, L. A. & Arntzen, C. J. Lateral mobility of the light-harvesting complex in chloroplast membranes controls excitation energy distribution in higher plants. *Arch. Biochem. Biophys.* **222**, 527–541 (1983).
41. Allen, J. F. Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain. *Trends Plant Sci.* **8**, 15–19 (2003).
42. Consoli, E., Croce, R., Dunlap, D. D. & Finzi, L. Diffusion of light-harvesting complex II in the thylakoid membranes. *EMBO Rep.* **6**, 782–786 (2005).
43. Kouril, R. *et al.* Structural characterization of a complex of photosystem I and light-harvesting complex II of *Arabidopsis thaliana*. *Biochemistry* **44**, 10935–10940 (2005).
44. Haldrup, A., Jensen, P. E., Lunde, C. & Scheller, H. V. Balance of power: a view of the mechanism of photosynthetic state transitions. *Trends Plant Sci.* **6**, 301–305 (2001).
45. Kargul, J. *et al.* Light-harvesting complex II protein CP29 binds to photosystem I of *Chlamydomonas reinhardtii* under State 2 conditions. *FEBS J.* **272**, 4797–4806 (2005).
46. Takahashi, H., Iwai, M., Takahashi, Y. & Minagawa, J. Identification of the mobile light-harvesting complex II polypeptides for state transitions in *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **103**, 477–482 (2006).
47. Trissl, H.-W. & Wilhelm, C. Why do thylakoid membranes from higher plants form grana stacks? *Trends Biochem. Sci.* **18**, 415–419 (1993).
48. Sener, M. K. *et al.* Evolution of the excitation transfer network in photosystem I from cyanobacteria to plants. *Biophys. J.* **89**, 1630–1642 (2005).
49. Amunts, A., Ben-Shem, A. & Nelson, N. Solving the structure of plant photosystem I—biochemistry is vital. *Photochem. Photobiol. Sci.* **4**, 1011–1015 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the ESRF for synchrotron beam time, and staff scientists of the ID 14, ID 29 and ID 23 station clusters for their assistance. We also thank F. Frolov and J. Hirsh for valuable guidance and advice in crystallography. This work was supported by The Israel Science Foundation.

Author Information Atomic coordinates and structure factor files were deposited in the Protein Data Bank under accession number 2O01. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.N. (nelson@post.tau.ac.il).

LETTERS

High-resolution subsurface water-ice distributions on Mars

Joshua L. Bandfield¹

Theoretical models indicate that water ice is stable in the shallow subsurface (depths of <1–2 m) of Mars at high latitudes^{1–7}. These models have been mainly supported by the observed presence of large concentrations of hydrogen detected by the Gamma Ray Spectrometer suite of instruments on the Mars Odyssey spacecraft^{8–10}. The models and measurements are consistent with a water-ice table that steadily increases in depth with decreasing latitude. More detailed modelling has predicted that the depth at which water ice is stable can be highly variable, owing to local surface heterogeneities such as rocks and slopes, and the thermal inertia of the ground cover^{11–13}. Measurements have, however, been limited to the footprint (several hundred kilometres) of the Gamma Ray Spectrometer suite, preventing the observations from documenting more detailed water-ice distributions. Here I show that by observing the seasonal temperature response of the martian surface with the Thermal Emission Imaging System on the Mars Odyssey spacecraft¹⁴, it is possible to observe such heterogeneities at subkilometre scale. These observations show significant regional and local water-ice depth variability, and, in some cases, support distributions in the subsurface predicted by atmospheric exchange and vapour diffusion models. The presence of water ice where it follows the depth of stability under current climatic conditions implies an active martian water cycle that responds to orbit-driven climate cycles^{15–17}. Several regions also have apparent deviations from the theoretical stability level, indicating that additional factors influence the ice-table depth. The high-resolution measurements show that the depth to the water-ice table is highly variable within the potential Phoenix spacecraft landing ellipses, and is likely to be variable at scales that may be sampled by the spacecraft.

The high concentrations of water ice inferred from Gamma Ray Spectrometer (GRS) measurements^{10,11} require that the water ice in the shallow subsurface on Mars be a mixture of water ice and regolith/rocky material. As a result, this subsurface water-ice/regolith mixture will have a thermal inertia similar to solid bedrock, that is, much higher than the more porous, dry, particulate regolith cover^{4,7,11}. Martian high latitudes have thermal properties consistent with an extensive high-thermal-inertia layer within a few centimetres of the surface, though extremely high thermal inertia values of $>1,000 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$, which are consistent with bedrock exposed at the surface, are rare¹⁸. The presence of this near-surface high-inertia layer with few surface exposures is consistent with the presence of a significant ice component that is insulated by several centimetres of ground cover. Although surfaces of increased thermal inertia are present at high latitudes, these surfaces commonly have thermal inertia values of only $400\text{--}500 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$.

The temperature response of a material is dependent on the intensity and period of the input energy cycle in addition to the thermophysical properties of the material itself. The thickness of the surface

layer that is influenced by the energy cycle (referred to as the skin depth) is proportional to the square root of the period of the cycle. Thus properties at a variety of depths can be determined from Phobos eclipse (roughly millimetre scales)¹⁹, diurnal (centimetre scales)^{20,21} and seasonal (decimetre to metre scales)^{22–25} surface temperature measurements. At high latitudes, the magnitude of the seasonal energy cycle is prominent, enhancing the surface temperature effect of a buried high-inertia water-ice-rich layer. At low latitudes, the lower seasonal variation in the energy cycle results in greatly diminished sensitivities to layers beyond diurnal skin depths.

Seasonal temperature measurements from the Thermal Emission Spectrometer (TES) on the Mars Global Surveyor spacecraft clearly indicate the presence of subsurface ice at high latitudes^{23–25}; however, the determination of the depth of this high-inertia ice layer is not precise. This is largely because variable surface properties (for example, surface frosts) within the 3 km resolution of TES and uncertain atmospheric properties can have significant effects on the modelled surface temperatures.

A more precise assessment of subsurface water-ice distributions can be obtained by measuring the seasonal change in relative temperatures between surfaces (Fig. 1). The high precision (in-flight measurements indicate a noise equivalent delta temperature of 1 K at 165 K) and high spatial resolution (100 m sampling) of night-time (local time of 03:30–05:30) Thermal Emission Imaging System (THEMIS) data are well suited for this application. By observing the change in relative temperatures between mid- to late-summer

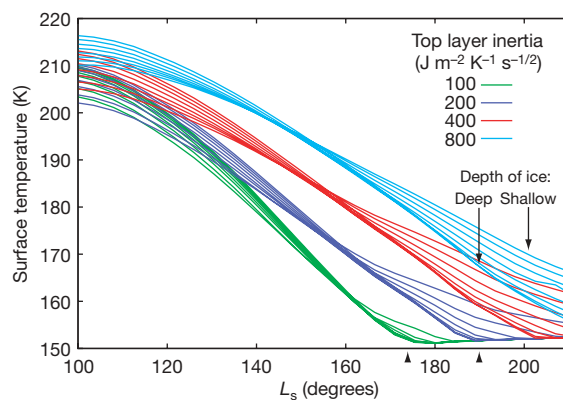


Figure 1 | Modelled temperatures for a variety of surface cover thermal inertias (top layer inertia) and ice-table depths, at 67.5°N. At this latitude, the high-inertia ice layer has little influence on surface temperatures near L_s (solar longitude) 160°, and primarily affects the rate of cooling throughout the late summer/early autumn season. Arrow points on the seasonal scale show the time of collection for the data used to generate Fig. 2. Temperatures are for a local time of 05:00, a surface albedo of 0.20 and a visible dust opacity of 0.30.

¹School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287-6305, USA.

and early autumn, it is possible to gain insight into the relative inertias of the surface layers and the relative depth to the high-inertia subsurface layer. The thermal inertia of the surface layer dominates the surface temperatures during the mid- to late-summer, and the change in relative temperatures between summer and early autumn is dominated by the depth of the ice layer (Fig. 1). Many uncertainties in modelling surface temperatures (for example, surface CO_2 frost, aerosol opacities, slope and albedo effects owing to low solar incidence) are avoided or largely cancel out using this method. Thus, surfaces that cool more quickly between summer and autumn have a deeper ice layer, and the slower cooling surfaces have a shallow ice layer.

The THEMIS data can be quantitatively supported by comparing the measured temperatures to those predicted by thermal models^{20,21} (the model used here was developed by H. H. Kieffer). Surface and subsurface temperatures are predicted, given proper environmental input parameters as well as a one-dimensional model of surface and subsurface properties containing layers of variable-inertia regolith cover and a high-inertia ice-bedrock layer. Although real world material distributions may be more complex, this model is useful for gaining initial insight into subsurface properties. In addition, the relatively simple nature of the temperature data used here prevents the recovery of additional information from a more complex

surface model. The model also predicts the maximum annual temperature of the water-ice layer. The temperature of water-ice stability is highly dependent on the role of atmospheric water vapour, and has been predicted to be 198 K with a mean annual water vapour column abundance of 10 precipitable micrometres (refs 7, 11).

Figure 2 displays seasonal temperature difference data from the Phoenix B region proposed landing site²⁶. Many of the slightly lower-albedo surfaces in the region display a higher night-time (and lower daytime) temperature throughout the summer season, indicative of a higher-inertia ground cover. These observations are supported by initial Mars Reconnaissance Orbiter High Resolution Imaging Science Experiment (HiRISE) images that indicate that these regions commonly have higher rock abundances²⁷ that would increase the average thermal inertia of the ground cover. The rocky surfaces change temperature more slowly on diurnal timescales and more quickly on seasonal timescales than the surrounding terrain through the early autumn season (Fig. 3). This temperature response is consistent with a deeper ice table in the low-albedo regions. The measured temperatures can be matched well in the low-albedo surfaces by a ground cover with a thermal inertia of $400 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ with an ice-bedrock layer at $>18 \text{ cm}$ (sensitivities are low at depths of greater than $\sim 20 \text{ cm}$). The seasonal temperatures of the higher-albedo surfaces can be well modelled by a ground cover thermal inertia of $200 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ and an ice-bedrock layer at 5 cm depth, consistent with lower-resolution measurements²⁵.

Similar data were processed for the Phoenix A region proposed landing site (shown in Supplementary Figs S1 and S2). In this case, regional differences in ice-table depth are clearly apparent (consistent with ice-table depths $> 9 \text{ cm}$), but the high degree of correlation with surface morphology at the subkilometre scale is not present.

In both examples, the modelled temperatures of the bedrock-ice layers never exceed 198 K, though the two regions apparently have a different character of subsurface ice distributions at this scale. A likely cause for this difference is the relatively variable surface-cover thermal inertia present at subkilometre scales in the Phoenix region B versus region A. The modelled surfaces are consistent with current orbital, atmospheric and thermophysical equilibrium conditions and free exchange with the martian atmosphere^{1–7,11–13}. This implies that ice deposition can occur in these regions by way of atmospheric vapour exchange.

Whereas near-surface ice stability generally decreases with increasing distance from the pole, local heterogeneities (such as slopes and variations in regolith cover) have a dominant effect on the local subsurface water-ice distributions^{11–13}. Far more variation in the ice-table depth can appear within a small section of the proposed

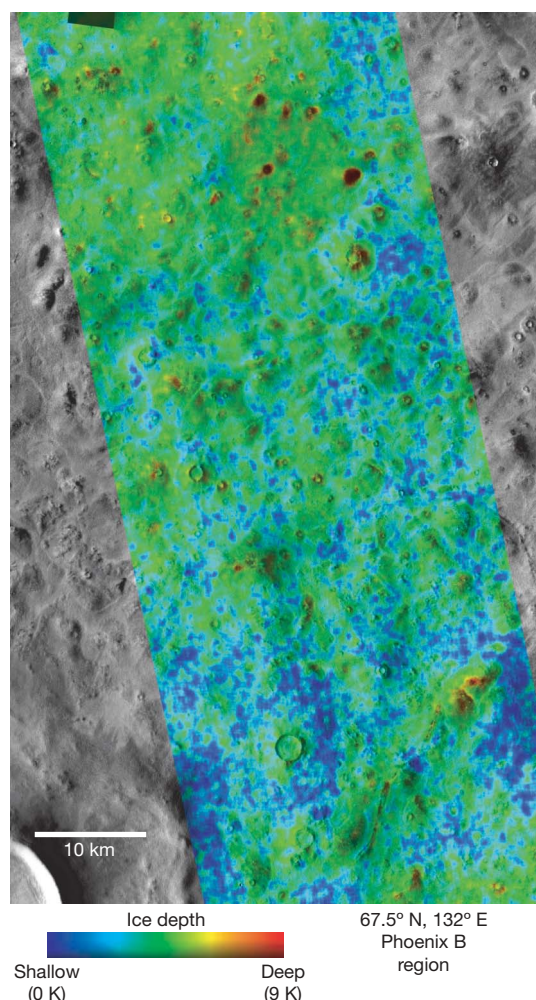


Figure 2 | Ice depth map centred near 67.5° N , 132° E (Phoenix B region proposed landing site). This map was produced by differencing the relative surface temperatures present in THEMIS night-time images I14393035 and I14705021, acquired at L_s 174° and 189° , respectively. Data are overlaid on a THEMIS visible image mosaic for morphological context. Modelled temperatures (Fig. 3) indicate that blue colours are consistent with water ice at 5 cm depth and red colours are consistent with water ice at $>18 \text{ cm}$.

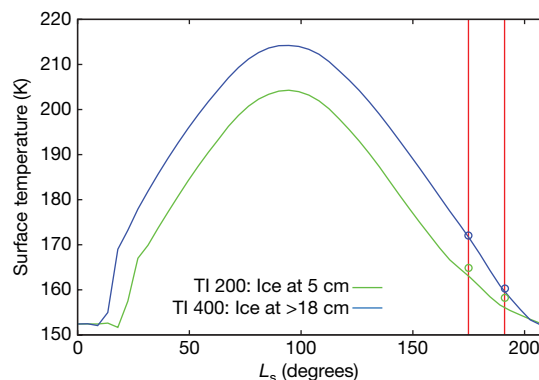


Figure 3 | Modelled temperatures at 05:00 for the surfaces shown in Fig. 2. Red vertical lines indicate the times of data acquisition, and circles are surface temperatures derived from THEMIS data. Both absolute temperatures and temperature differences between the shallow (green curve; top layer thermal inertia (TI) of $200 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$) and deep (blue curve; top layer thermal inertia of $400 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$) ice surfaces shown in Fig. 2 are matched well (within 5 K absolute and 2 K relative) by the model.

Phoenix landing site regions than predicted by the theoretical stabilities throughout the region based on low-resolution data sets. In addition, high-resolution images show clear indications of surface heterogeneity, implying that a highly variable ice table can be present at even smaller spatial scales than presented here.

Figure 4 displays seasonal temperature difference data in the southern hemisphere near Melea Planum. This region is marked by a smooth (at 100 m scales) layer that appears to be a remnant of a more extensive unit, rough hummocky terrain, and sand deposits that collect in local depressions. The surface thermophysical properties display little correlation with surface morphology, except where sand dune forms are present. The sand deposits have the largest amount of seasonal cooling, consistent with a relatively deep ice table compared with the surrounding terrain. Surface temperatures can be fitted with a modelled ground cover thermal inertia of $220 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ with ice present at $>19 \text{ cm}$ for the sand deposits (shown in Supplementary Fig. S3). The surrounding terrain surface temperatures are consistent with a ground cover inertia of $150 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ with ice present at 1 cm depth. Several areas displayed in Fig. 4 have an even lower degree of seasonal cooling, indicating an even shallower ice table.

The sand deposits are consistent with either an ice table limited by the 198 K temperature constraint or no ice table at all. A well-sorted

sand would be highly permeable and permit free exchange of the subsurface with the atmosphere, limiting the depth of water-ice stability to where this temperature is not exceeded. In addition, if the sand is mobile, any shallow ice table would be eventually exposed to non-equilibrium temperature conditions close to the surface. Outside the sand deposits, the ice table is apparently significantly shallower than predicted by the 198 K temperature of ice stability. This could be due to either an effective seal of the ground cover that would prevent effective atmospheric exchange and/or higher local annual-mean water vapour abundances that would raise the temperature of water-ice stability. The surface temperature data could not be reasonably fitted by a model where the 198 K subsurface ice temperature constraint is preserved.

The seasonal temperature data indicate clear subsurface water-ice heterogeneity, as well as regional differences in the pattern of ice distributions. As has been noted^{11–13}, the nature of the ground cover has a dominant effect on the depth of stability of the water-ice table. Higher-inertia materials will have greater temperatures at depth because of their higher thermal conductivity and because their cooler daytime temperatures do not radiate the heat away as effectively as low-inertia materials. The THEMIS seasonal temperature data in the examples shown here largely support the depths of water ice predicted by theoretical models^{1–7,11–13}. As also shown by the GRS measurements, the presence of ice at its current stability limits may indicate that large changes in stability predicted to accompany orbital variations drive an active water cycle between the poles, regolith and atmosphere^{5,15,16}.

There appear to be instances where the presence of ice does not, however, follow its limit of stability. Because the sensitivity of the temperature data to the presence of a high-inertia layer is limited to a couple of decimetres, regions where ice is predicted to be greater than $\sim 20 \text{ cm}$ may indeed have no subsurface ice at all. In addition, the presence of the shallow ice in the southern hemisphere example (Fig. 4) appears to clearly violate the assumed ice stability conditions. The detailed distribution of water ice in martian high latitudes appears to be complex, with controlling factors beyond those predicted by vapour diffusion models. As noted in ref. 28, there are reasonable geologic situations that would prevent or considerably slow the exchange of water between atmosphere and regolith. This supports the notion that shallow water ice may be present on Mars at lower latitudes than predicted by vapour diffusion models.

The regions presented here were limited by the fortuitous acquisition of at least two images of the same surface during specific seasons and local times. More directed targeting would allow a number of regions to be studied, so as to assemble a more complete picture of the martian subsurface water-ice distributions.

METHODS SUMMARY

THEMIS data were calibrated using methods described elsewhere^{14,29}. An additional correction was made for potential focal plane temperature drift between acquisition of the image and calibration data. Because of the temperature stability of the martian atmosphere during these seasons ($<1\text{--}2 \text{ K}$ inter-annual variation³⁰), the resulting calibrated THEMIS data are accurate to within 5 K at 160 K.

Data were converted to brightness temperature from the $12.6 \mu\text{m}$ THEMIS band 9. Because of the relatively low atmospheric opacity, low surface-atmosphere temperature contrast, and high surface emissivity for the observations used in this study, brightness temperatures are within 1–2 K of surface kinetic temperature. Temperature images taken over a single surface at two seasons were subtracted from each other in order to isolate the relative temperature changes between surfaces. This subtraction also cancels the absolute temperature uncertainties discussed above.

Modelled temperatures were fitted to the measured data by manually adjusting regolith cover thermal inertia and depth to ice-bedrock layer until the predicted surface temperatures fall within the uncertainties of the measured THEMIS surface temperatures. Although it is possible to automate a fitting routine^{24,25}, it is difficult to adjust the weighting of the factors to produce an intelligent convergence. In addition, uncertainties in the model, and in surface and atmospheric properties fed into the model, preclude a more accurate

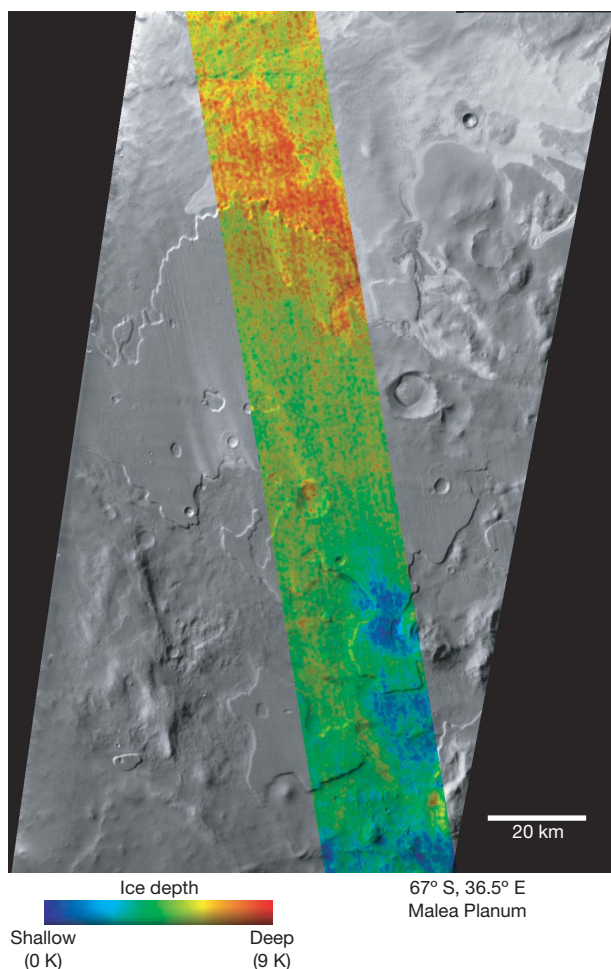


Figure 4 | Ice depth map centred near 67° S, 36.5° E. This map was produced by differencing the relative surface temperatures present in THEMIS night-time images I17904012 and I18216009, acquired at L_s 347° and 0°, respectively. Data are overlaid on a THEMIS daytime infrared image mosaic for morphological context. Low-albedo duneforms appear bright because of their warmer daytime temperatures. Modelled temperatures (Supplementary Fig. S3) indicate that blue colours are consistent with water ice at $<1 \text{ cm}$ depth, and red colours are consistent with water ice at $>19 \text{ cm}$.

determination than what can be produced manually. As a result, the model fits are used as an example of consistency with surface temperature measurements, but uncertainties in ice-table depth and regolith cover inertia are poorly understood. Sensitivity to the depth of the ice-bedrock layer is greatest where the layer is shallow, the surface cover thermal inertia is low, and the latitude is 40–90°.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 December 2006; accepted 26 March 2007.

- Leighton, R. R. & Murray, B. C. Behavior of carbon dioxide and other volatiles on Mars. *Science* **153**, 136–144 (1966).
- Fanale, F. P., Salvail, J. R., Zent, A. P. & Postawko, S. E. Global distribution and migration of subsurface ice on Mars. *Icarus* **67**, 1–18 (1986).
- Zent, A. P., Fanale, F. P., Salvail, J. R. & Postawko, S. E. Distribution and state of H₂O in the high-latitude shallow subsurface of Mars. *Icarus* **67**, 19–36 (1986).
- Paige, D. A. The thermal stability of near-surface ground ice on Mars. *Nature* **356**, 43–45 (1992).
- Mellon, M. T. & Jakosky, B. M. The distribution and behavior of Martian ground ice during past and present epochs. *J. Geophys. Res.* **100**, 11781–11799 (1995).
- Jakosky, B. M., Zent, A. P. & Zurek, R. W. The Mars water cycle: Determining the role of exchange with the regolith. *Icarus* **130**, 87–95 (1997).
- Boynton, W. V. *et al.* Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. *Science* **297**, 81–85 (2002).
- Schorghofer, N. & Aharonson, O. Stability and exchange of subsurface ice on Mars. *J. Geophys. Res.* **110**, E5003, doi:10.1029/2004JE002350 (2005).
- Feldman, W. C. *et al.* Global distribution of near-surface hydrogen on Mars. *J. Geophys. Res.* **109**, E9006, doi:10.1029/2003JE002160 (2004).
- Mitrofanov, I. G. *et al.* Soil water content on Mars as estimated from neutron measurements by the HEND instrument onboard the 2001 Mars Odyssey Spacecraft. *Solar Syst. Res.* **38**, 253–257 (2004).
- Mellon, M. T., Feldman, W. C. & Prettyman, T. H. The presence and stability of ground ice in the southern hemisphere of Mars. *Icarus* **169**, 324–340 (2004).
- Sizemore, H. G. & Mellon, M. T. Effects of soil heterogeneity on martian ground-ice stability and orbital estimates of ice table depth. *Icarus* **185**, 358–369 (2006).
- Aharonson, O. & Schorghofer, N. Subsurface ice on Mars with rough topography. *J. Geophys. Res.* **111**, E11007, doi:10.1029/2005JE002636 (2006).
- Christensen, P. R. *et al.* The Thermal Emission Imaging System (THEMIS) for the Mars 2001 Odyssey Mission. *Space Sci. Rev.* **110**, 85–130 (2004).
- Haberle, R. M. & Jakosky, B. M. Sublimation and transport of water from the north residual polar cap on Mars. *J. Geophys. Res.* **95**, 1423–1437 (1990).
- Jakosky, B. M., Henderson, B. G. & Mellon, M. T. Chaotic obliquity and the nature of the Martian climate. *J. Geophys. Res.* **100**, 1579–1584 (1995).
- Head, J., Mustard, J., Kreslavsky, M., Milliken, R. & Marchant, D. Recent ice ages on Mars. *Nature* **426**, 797–802 (2003).
- Edwards, C. S., Bandfield, J. L., Christensen, P. R. & Fergason, R. L. Global distribution of bedrock on Mars using THEMIS high resolution thermal inertia. *Eos* **158** (Fall Mtg), abstr. P21C–0158 (2005).
- Betts, B. H., Murray, B. C. & Svitek, T. Thermal inertias in the upper millimeters of the Martian surface derived using Phobos' shadow. *J. Geophys. Res.* **100**, 5285–5296 (1995).
- Kieffer, H. H. *et al.* Thermal and albedo mapping of Mars during the Viking primary mission. *J. Geophys. Res.* **82**, 4249–4291 (1977).
- Mellon, M. T., Jakosky, B. M., Kieffer, H. H. & Christensen, P. R. High-resolution thermal inertia mapping from the Mars Global Surveyor Thermal Emission Spectrometer. *Icarus* **148**, 437–455 (2000).
- Paige, D. A., Bachman, J. E. & Keegan, K. D. Thermal and albedo mapping of the polar regions of Mars using Viking thermal mapper observations: 1. North polar region. *J. Geophys. Res.* **99**, 25959–25991 (1994).
- Titus, T. N., Kieffer, H. H. & Christensen, P. R. Exposed water ice discovered near the south pole of Mars. *Science* **299**, 1048–1051 (2003).
- Armstrong, J. C., Titus, T. N. & Kieffer, H. H. Evidence for subsurface water ice in Korolev crater, Mars. *Icarus* **174**, 360–372 (2005).
- Titus, T. N., Prettyman, T. P. & Colaprete, A. Thermal characterization of the three proposed Phoenix landing sites. *Lunar. Planet. Sci. Conf.* **37**, abstr. 2161 (2006).
- Arvidson, R. E. *et al.* Overview of Mars exploration program 2007 Phoenix mission landing site selection. *Lunar. Planet. Sci. Conf.* **37**, abstr. 1328 (2006).
- Polygonal terrain in the northern plains. (http://marsweb.nas.nasa.gov/HIRISE/hirise_images/TRA/TRA_000828_2495/) (2006).
- Jakosky, B. M. *et al.* Mars low-latitude neutron distribution: Possible remnant near-surface water ice and a mechanism for its recent emplacement. *Icarus* **175**, 58–67 (2005).
- Bandfield, J. L., Rogers, D., Smith, M. D. & Christensen, P. R. Atmospheric correction and surface spectral unit mapping using Thermal Emission Imaging System data. *J. Geophys. Res.* **109**, E10008, doi:10.1029/2004JE002289 (2004).
- Smith, M. D. Interannual variability in TES atmospheric observations of Mars during 1999–2003. *Icarus* **167**, 148–165 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Thanks to P. Christensen, R. Fergason, H. Kieffer, C. Edwards, R. Luk, K. Bender, and J. Hill for data processing and targeting help and discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (joshband@asu.edu).

METHODS

THEMIS data were calibrated using methods described elsewhere^{14,29}. An additional correction was performed by adjusting the data number of all bands until the measured radiance in THEMIS band 10 matched that predicted by TES data within the 15 μm CO_2 fundamental. This corrects for potential focal plane temperature drift between acquisition of the image and calibration data. Because of the temperature stability of the martian atmosphere during these seasons (<1 – 2 K inter-annual variation³⁰), the resulting calibrated THEMIS data are accurate to within 5 K at 160 K.

Data were converted to brightness temperature from the 12.6 μm THEMIS band 9 using a look-up table of Planck radiances convolved with the THEMIS band 9 filter response. Because of the relatively low atmospheric opacity ($\tau < 0.05$), low surface–atmosphere temperature contrast (<20 K), and high surface emissivity ($\varepsilon > 0.98$) at 12.6 μm for the observations used in this study, brightness temperatures are within 1–2 K of surface kinetic temperature.

THEMIS images were selected by the following criteria: (1) all image incidence angles are $>90^\circ$. This ensures that albedo and surface slopes have little effect on the surface temperatures in the pre-dawn images; (2) all images have average band 9 brightness temperatures >155 K and no regions contain significant areas at CO_2 condensation temperatures (~ 148 K); (3) all images were acquired in the late summer or early autumn season, and each image pair was acquired in the same year at a similar local time with a separation of at least $10^\circ L_s$. The image pairs used for this work were I14393035–I14705021 (Fig. 2), I17904012–I18216009 (Fig. 4), and I14401005–I14713023 (Supplementary Fig. S1).

To produce the depth maps, the two temperature images were subtracted from each other and the level was adjusted to make the surface with the least amount of cooling equal zero. This isolates the relative temperature changes between surfaces and cancels the absolute temperature uncertainties discussed above. On the basis of the model results displayed in Fig. 1, this subtraction also provides the relative rate of seasonal cooling, which is largely independent of surface cover thermal inertia and highly dependent on the burial depth of the subsurface water ice.

Modelled temperatures were fitted to the measured data by manually adjusting regolith cover thermal inertia and depth to ice–bedrock layer until the predicted surface temperatures fall within the uncertainties of the measured THEMIS surface temperatures. The ice–bedrock layer used the properties for density, thermal conductivity, and heat capacity listed in ref. 11. Although it is possible to automate a fitting routine^{23,24}, it is difficult to adjust the weighting of the factors to produce an intelligent convergence. In addition, uncertainties in the model and surface and atmospheric properties fed into the model preclude a more accurate determination than what can be produced manually. As a result, the model fits are used as an example of consistency with surface temperature measurements, but uncertainties in ice-table depth and regolith cover inertia are poorly understood. However, dependence on several factors such as heterogeneous surfaces (lateral mixtures of rocks and fine particulates), surface albedo, and atmospheric dust properties were tested and found to have little effect on the predicted temperatures at the local time and season of image acquisition for the data used in this study. For example, a surface composed of rocks and soil will simply raise the thermal inertia of the surface cover relative to a homogeneous soil and have little effect on the modelled rate of seasonal cooling.

The sensitivity of the measurements to the presence and depth of water ice is dependent on three main factors; the ground cover thermal inertia, the depth of the ice, and latitude. The thermal model used in this work divides layers on the basis of units of skin depth rather than absolute depth in order to achieve similar levels of temperature precision for a variety of thermal inertia values. Because the skin depth is proportional to the thermal inertia, the model will have twice the vertical resolution for a surface cover thermal inertia of 200 versus $400 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$, for example. The surface temperature influence of a high-inertia ice-rich layer is reduced by a factor of ~ 2 – 3 with each doubling of its depth. For example, at 65°N with a surface cover thermal inertia of $220 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$, an ice layer at 5 cm will have up to a 7.5 K effect during the autumn season versus 4 K at 10 cm and 1.3 K at 20 cm. Lastly, the latitude influences the magnitude of seasonal variation in energy input to the surface. A surface cover thermal inertia of $220 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ and an ice layer at 5 cm will have up to an 8–10 K effect during the autumn season at latitudes of 40 – 90° versus <3 K at latitudes of $<30^\circ$.

The combination of these effects greatly reduces the uncertainty in water-ice depths where the depths are shallow, the surface cover thermal inertia is low, and the latitude is $>40^\circ$. For example, with a surface cover thermal inertia of $150 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ and a water-ice depth of 1 cm (similar to surfaces shown in Fig. 4), a change to 0.5 cm in depth will cause up to a 3 K difference in surface temperature during the autumn season. With a surface cover thermal inertia of

$300 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ and a water-ice depth of 14 cm, a change to 20 cm in depth is required to cause the same 3 K difference in surface temperature.

LETTERS

Direct measurement of antiferromagnetic domain fluctuations

O. G. Shpyrko¹, E. D. Isaacs^{1,3}, J. M. Logan³, Yejun Feng³, G. Aeppli⁴, R. Jaramillo³, H. C. Kim³, T. F. Rosenbaum³, P. Zschack², M. Sprung², S. Narayanan² & A. R. Sandy²

Measurements of magnetic noise emanating from ferromagnets owing to domain motion were first carried out nearly 100 years ago¹, and have underpinned much science and technology^{2,3}. Antiferromagnets, which carry no net external magnetic dipole moment, yet have a periodic arrangement of the electron spins extending over macroscopic distances, should also display magnetic noise. However, this must be sampled at spatial wavelengths of the order of several interatomic spacings, rather than the macroscopic scales characteristic of ferromagnets. Here we present a direct measurement of the fluctuations in the nanometre-scale superstructure of spin- and charge-density waves associated with antiferromagnetism in elemental chromium. The technique used is X-ray photon correlation spectroscopy, where coherent X-ray diffraction produces a speckle pattern that serves as a ‘fingerprint’ of a particular magnetic domain configuration. The temporal evolution of the patterns corresponds to domain walls advancing and retreating over micrometre distances. This work demonstrates a useful measurement tool for antiferromagnetic domain wall engineering, but also reveals a fundamental finding about spin dynamics in the simplest antiferromagnet: although the domain wall motion is thermally activated at temperatures above 100 K, it is not so at lower temperatures, and indeed has a rate that saturates at a finite value—consistent with quantum fluctuations—on cooling below 40 K.

Because of scientific and technical interest in ferromagnetic domains, there has been substantial, long-standing work on magnetic noise in ferromagnets as a direct witness of domain motion. As antiferromagnets begin to find applications themselves—for example, as pinning layers in spintronics—there is a need for measurements of the noise associated with moving antiferromagnetic domains. Antiferromagnetic domain dynamics are also important because they are implicated in basic problems in condensed matter physics, such as high temperature superconductivity and ‘heavy’ fermions. Neutrons are an excellent non-local probe of antiferromagnetism and its dynamics⁴. However, a direct local probe of mesoscopic antiferromagnetic domain dynamics has not been hitherto available; this is because the magnetic dipole moments for antiferromagnets vanish on the scale of a nanometre, rendering the domain fluctuations responsible for noise essentially invisible to the direct magnetometer probes (for example, superconducting interference devices) that have been so successful for ferromagnets⁵.

Chromium is a body-centred cubic (b.c.c.) metal with an antiferromagnetic state nearly described by the simple rule that the electrons surrounding each Cr atom have magnetization opposite to those on the nearest neighbour Cr atoms. What actually occurs is sinusoidal modulation of this elementary magnetic structure—called a spin

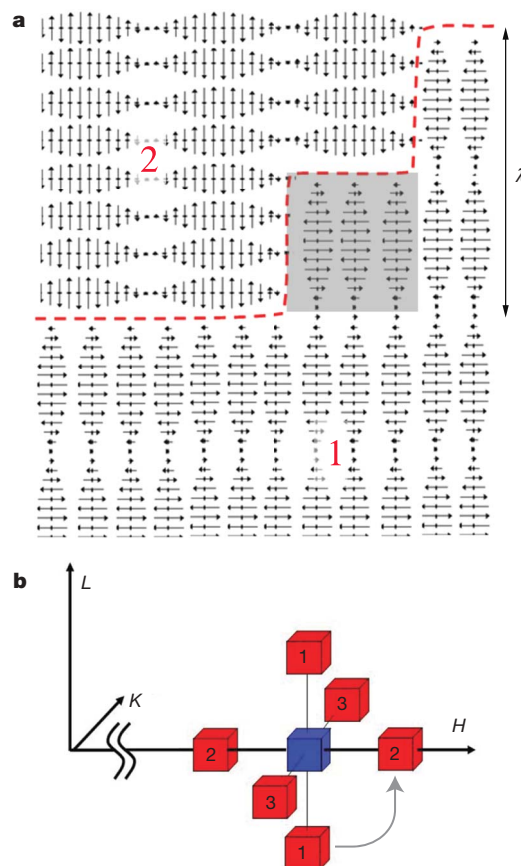


Figure 1 | Spin-density wave (SDW) domain wall in chromium. **a**, Schematic representation of the domain wall (red dashed line) separating two regions with perpendicular orientations of transverse SDW. The domain wall is shown to propagate along the weak nodal planes—where local magnetization approaches zero. Shaded region represents an elementary domain unit with volume $(\lambda/2)^3$, where λ is the SDW period, that can be thought of as a magnetic quantum dot in a cubic lattice of similar quantum dots. **b**, Reciprocal space configuration of the lattice [200] Bragg peak (blue) and six surrounding charge-density wave (CDW) satellites (red). H , K and L define fundamental axes of the reciprocal cubic lattice. Domains marked as 1 and 2 in **a** contribute to pairs of satellites marked 1 and 2 in **b**, respectively. A 90° rotation of SDW propagation vector within the shaded elementary volume of domain 1 would realign spins with domain 2, resulting in a shift of the domain wall and a transfer of scattering intensity from satellite pairs 1 to 2, marked with an arrow in **b**.

¹Center for Nanoscale Materials, ²Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois 60439, USA. ³James Franck Institute and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA. ⁴London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.

density wave (SDW), with wavelength $\lambda = 6\text{--}8\text{ nm}$ —along one of the three equivalent cubic (100) directions. A single crystal chromium sample cooled below the Néel temperature, $T_N = 311\text{ K}$, spontaneously breaks (see Fig. 1) into three types of magnetic domains characterized by the three different choices for the SDW propagation direction (for a review of SDWs in Cr, see ref. 6). The SDW is accompanied by a charge density wave (CDW), a combination of both itinerant and ionic charge modulation.

X-ray microdiffraction reveals that the typical size of the SDW domains in bulk Cr samples is of the order of $1\text{--}30\text{ }\mu\text{m}$ (ref. 7). Fluctuations of domain walls at fixed temperature have been studied via random electrical telegraph noise in thin Cr films for temperatures above 140 K (ref. 8). Even though the measurements were done for mesoscopic samples, the effects on the electrical resistance (R) of the switching dynamics were small ($\delta R/R \approx 10^{-5}$) and the interpretation difficult because R is an indirect probe of the underlying SDW and CDW order.

We report the first direct observations of domain wall fluctuations in bulk Cr using X-ray photon correlation spectroscopy (XPCS), which overcomes the limitations of the classic bulk and laser probes in that it accesses directly the short wavelength structure associated with the SDW. A coherent beam illuminating a partially ordered system (in our case consisting of SDW/CDW domains) produces an interference pattern, also known as speckle^{9,10}. Owing to the high sensitivity of speckle to minute changes in domain wall configuration, the time variation of the speckle pattern directly reveals the dynamics of domain structure. Figure 2a is a diagram of the experimental configuration, and Fig. 2b shows a speckle pattern of the (200) Bragg peak for the b.c.c. Cr lattice. Interference fringes arising from partial coherence of the X-ray beam are clearly seen in the image, as well as in the line scans (Fig. 2c). Incoherent diffraction would produce the gaussian-like profile represented by the black line in Fig. 2c.

The lattice Bragg speckle pattern is static over 5 h, indicating the high level of stability for our instrumentation and the sample.

We turn next to the speckle pattern for the $[2-\delta, 0, 0]$ CDW superlattice reflection, displayed for 17 K at a variety of times in Fig. 3b. The patterns in subsequent frames, separated by $1,000\text{ s}$, grow increasingly dissimilar for longer time lags—patterns within frames collected more than $3,000\text{ s}$ apart appear completely uncorrelated. Thus, the CDW speckle evolves with a characteristic time of a few thousand seconds or less, much shorter than the $>20,000\text{ s}$ relaxation time for the b.c.c. Bragg speckle of Fig. 2c. This indicates that the changes in the CDW speckle are indeed due to changes in the magnetic domain configuration, rather than some experimental artefact. For example, drift of the X-ray beam or the cryostat, motion of crystalline defects within the Cr sample, or any other effect not related to magnetic domain dynamics would inevitably cause changes in both the CDW and (200) Bragg speckle.

The spatial sensitivity of the speckle to domain motion is described by two distinct lengths: the first is $1/\Delta Q \approx 100\text{ }\text{\AA}$, where $\Delta Q = 10^{-2}\text{ }\text{\AA}^{-1}$ is the total size of visible speckle pattern in reciprocal space (see Figs 2b, c and 3b) and represents the minimum size of domains with a visible impact on the speckle pattern. The second is the domain wall displacement necessary to produce a speckle pattern that is highly dissimilar (or uncorrelated) to the original one. A combination of X-ray microdiffraction images of domain configurations and speckle simulations indicate that this second length is $1\text{ }\mu\text{m}$ (see Methods and Supplementary Information).

Beyond revealing that domain walls are moving by distances of the order of $1\text{ }\mu\text{m}$, the data provide several other important quantities. For example, we can evaluate the autocorrelation function, $g_2(t)$:

$$g_2(t) = \frac{\langle I(\tau)I(\tau+t) \rangle_\tau}{\langle I(\tau) \rangle_\tau^2} = 1 + A|F(\mathbf{Q}, t)|^2 \quad (1)$$

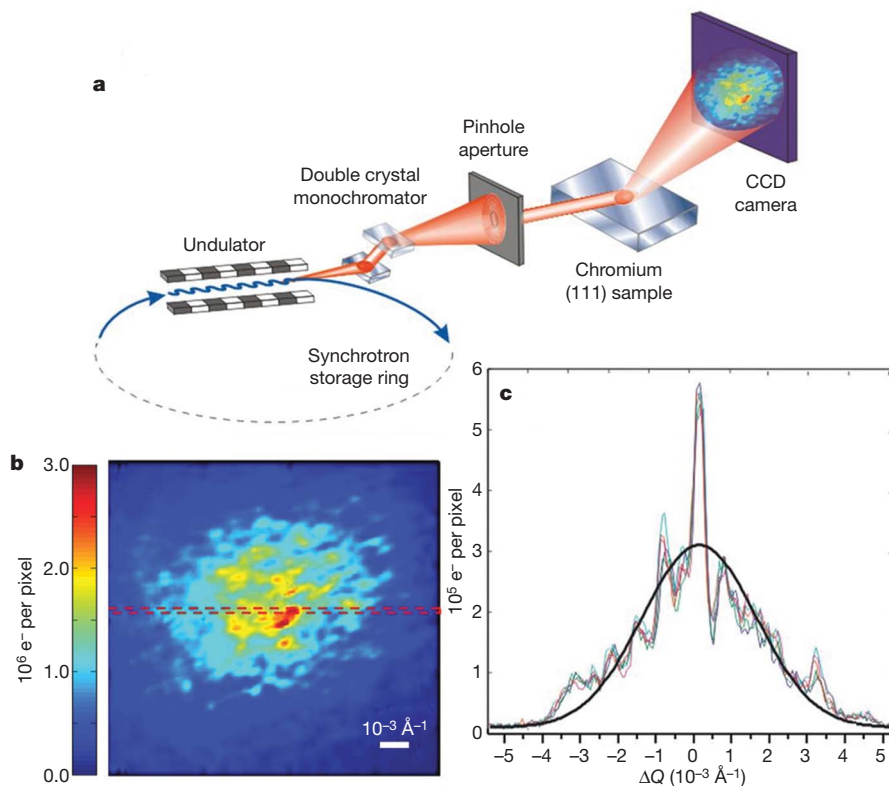


Figure 2 | X-ray speckle measurements. **a**, Diagram of the experimental set-up. **b**, CCD image of the X-ray speckle observed for the [200] lattice Bragg reflection. **c**, Intensity distribution for a line scan across the region between the dashed lines in **b**. Five differently coloured and nearly identical

lines represent line scans of the portion of speckle pattern shown between the red dashed lines in **b**, taken one hour apart. The black line is a simulated statistically averaged gaussian profile, expected for a completely incoherent beam.

where $I(\tau)$ and $I(\tau + t)$ are the intensities in a given pixel for frames taken at times τ and $\tau + t$ respectively, $F(\mathbf{Q}, t)$ is the intermediate scattering function, A describes the beam coherence^{9,10}, and the averaging is performed over times τ and pixels corresponding to wavevector \mathbf{Q} . Figure 3a shows $|F(\mathbf{Q}, t)|^2$ for several temperatures calculated from the CDW speckle at $\mathbf{Q} = [2-2\delta, 0, 0]$. For large time delays the speckle patterns become uncorrelated, resulting in $g_2(t) = 1$, corresponding to $|F(\mathbf{Q}, t)|^2 = 0$. The dynamics are strongly temperature-dependent: on cooling, the domain fluctuation times increase by nearly two decades. Surprisingly, below 40 K the times remain finite, rather than diverging as expected for thermally driven fluctuations.

Two distinct fluctuation timescales are visible in most data sets presented in Fig. 3a. The measured $|F(\mathbf{Q}, t)|^2$ was therefore modelled by a double exponential form:

$$|F(\mathbf{Q}, t)|^2 = a \exp\left(-\left(t/\tau_F\right)^\beta\right) + (1-a) \exp\left(-\left(t/\tau_S\right)^\beta\right) \quad (2)$$

where τ_F and τ_S represent the characteristic fast and slow fluctuation timescales, respectively. A small value of $a = 0.03-0.10$ indicates that the time dependence of $|F(\mathbf{Q}, t)|^2$ is mainly due to slow fluctuations. The value of the stretching exponent β was found to be greater than 1, manifested by the 'compressed' shape of the $|F(\mathbf{Q}, t)|^2$. Compressed exponential relaxation has been observed for a variety of soft matter systems undergoing 'jamming' transitions, which result in arrested, solid-like collective dynamics¹¹⁻¹³ with $\beta > 1$, instead of liquid-like fluctuations with $\beta \leq 1$. Extended to our system, this points to elastically coupled dynamics between blocks of spins, similar to elastic collective depinning dynamics observed in CDW conductors¹⁴, an observation also consistent with the weakly pinned nature of SDW/CDW domains¹⁵⁻¹⁷. Furthermore, the fit value of β at $T < 100$ K is approximately 1.5 (Fig. 4, left inset), a universal value for dynamics of soft condensed matter systems in a jammed state¹⁸.

Figure 4 shows the T -dependence of the slow relaxation times τ_S obtained from fits to autocorrelation functions in Fig. 3a. The 20% uncertainty in fitting parameters τ_S arises primarily from counting

statistics of the autocorrelation function $g_2(t)$ (see Supplementary Information). Standard thermal activation ($\tau_S^{-1} = f_0 \exp(-\Delta E/k_B T)$, blue line) with a single attempt frequency f_0 and activation barrier $\Delta E/k_B = 240 \pm 50$ K accounts for the data at high T . The thermal picture fails spectacularly at low temperature for $T < 40$ K, and a switching mechanism which is temperature-independent in this range is required. The simplest possibility is that switching between low-energy domain wall configurations occurs via quantum tunnelling, rather than classical thermal activation (Fig. 4, right inset). A fit to the data that combines a thermally activated model and a quantum tunnelling contribution represented by a temperature-independent residence time τ_{QT} , $\tau_S^{-1} = \tau_{QT}^{-1} + \tau_R^{-1} \exp(-\Delta E/k_B T)$ is shown by the red solid line in Fig. 4 for $\tau_{QT} = 5,000$ s and $\tau_R = 15$ s (confidence limits obtained from the fits are $\tau_{QT} = 5,000 \pm 1,000$ s and $\tau_R = 4-60$ s). In analogy with α and β relaxation observed in glasses, supercooled liquids and jammed soft matter systems, faster fluctuations represent local relaxation, while slower fluctuations are due to collective relaxation modes.

The relaxation times observed here are similar to those associated with magnetization switching in ferromagnets first observed by Barkhausen¹ and studied since then in systems ranging from bulk materials to magnetic molecules^{5,19-22}. Antiferromagnets have more complex order than ferromagnets because they break translation as well as spin rotation invariance, which has forced us to formulate a very crude physical picture in order to understand our data at a semiquantitative level. We start with the realization that to minimize the very large exchange energy (> 0.4 eV)^{23,24} associated with domain walls, it is clearly advantageous for the domain walls to lie on the nodal planes²⁵ (where the spin polarization vanishes). Such an assumption is further supported by the previously observed preference for the formation of SDW nodes at Fe/Cr interfaces²⁶. This implies that the fundamental switching unit (grey shaded region in Fig. 1a) has volume $V_S \approx (\lambda/2)^3$, where λ is the underlying period of the SDW. In the simplest gaussian model where underlying units are switching randomly at typical times τ_U , we would conclude that the switching time for a volume of $V = 1 \mu\text{m}^3$ would be $(V/V_S)^2 \tau_U$. Using our experimental value for the attempt frequency τ_R^{-1} , we therefore

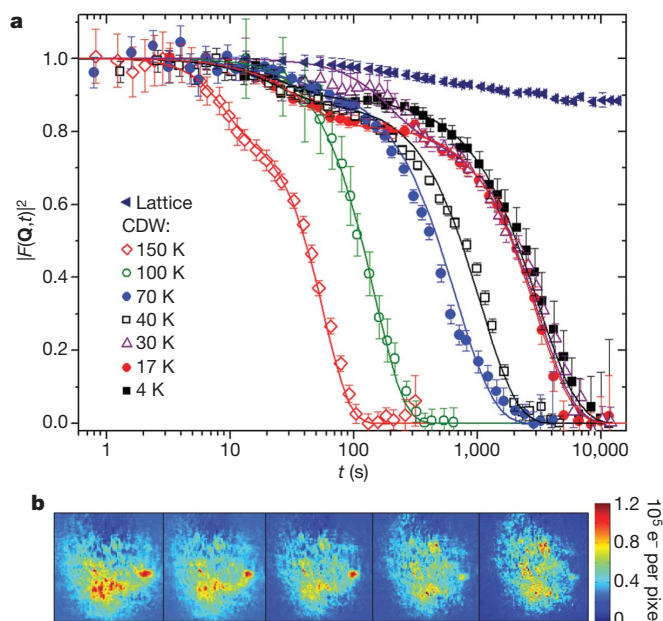


Figure 3 | Autocorrelation of speckle images. **a**, Intensity autocorrelation data for the [200] lattice Bragg peak, as well as for the CDW superlattice $[2-2\delta, 0, 0]$ peak, at temperatures T (in K) 150, 100, 70, 40, 30, 17 and 4. Error bars, standard deviation calculated for a region of interest within the CCD typically containing $n = 4 \times 10^4$ pixels. Two distinct timescales are clearly present in the CDW autocorrelation function. Solid lines represent theoretical fits to the data. See text for further details. **b**, Time sequence of CDW speckle pattern evolution at 17 K. Subsequent images are taken 1,000 s apart; each image is $10^{-2} \text{ \AA}^{-1} \times 10^{-2} \text{ \AA}^{-1}$.

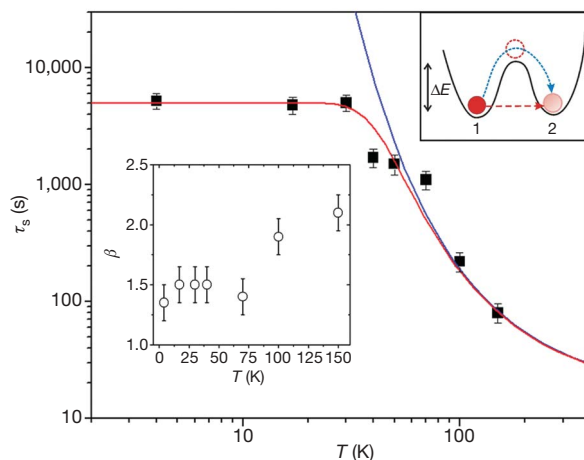


Figure 4 | Temperature-dependent domain wall dynamics. Main plot, characteristic slow fluctuation timescale, τ_S , obtained from fits to autocorrelation function data shown in Fig. 3a, compared to a classical Arrhenius model (blue line) and a model that also includes a temperature-independent switching rate term (red line). Right inset, potential energy surface including thermally activated (blue dashed line) and quantum tunnelling (red dashed line) mechanisms of the transition between the two low energy domain configurations 1 and 2 (see Fig. 1 for an example involving elementary switching volume) separated by energy barrier ΔE . Left inset, values of stretching exponent β for various temperatures. Error bars for τ_S and β represent the range of parameters obtained from fitting the autocorrelation data in Fig. 3.

obtain $\tau_U^{-1} = 36 \text{ THz} = 140 \text{ meV}$ as the attempt frequency for rotating an entire unit. This is an electronic energy scale, and could therefore be derived from the hopping of electrons across the domain wall; such electrons (also important in electrical noise measurements⁸) are, after all, responsible for the current fluctuations that sample the possibility of rotating the Fermi surface of a 'quantum dot' with the fundamental unit volume V_S . The tunnelling degree of freedom, which involves rotating the Fermi surface and its proxy, the \mathbf{Q} vector, can be considered a spherical rotor in a potential with minima along the cubic x , y and z directions. Our measured barrier height ΔE is not far from the Néel temperature, which allows us to think of the antiferromagnetic phase transition as occurring when Q -domain walls become unstable to thermal fluctuations. The barriers between the potential minima could also be identified by magnetic neutron scattering, such as that performed for $\text{Cr}_{0.95}\text{V}_{0.05}$ (ref. 27). This underscores the fundamental complementary nature of the XPCS measurements of the dynamics as measured in the time domain, and dynamics measured by neutrons in the frequency domain.

In the simplest Wentzel-Kramers-Brillouin (WKB) approximation (see, for example, ref. 22), the dimensionless ratio τ_R/τ_{QT} is equal to $\exp(-S/\hbar)$, where $S = \sqrt{I\Delta E}$ is the tunnelling action and I is the moment of inertia of a single rotor. Because the underlying attempt frequencies and their rescaling to account for observable effects in the X-ray experiment are the same for both incoherent quantum and classical processes, all of the detail—invoking multiple rotors—of the last paragraph drops out, and S characterizes a single rotor. We can therefore calculate the moment of inertia I of the quantum rotor using the measured parameters τ_{QT} , τ_R and the barrier height $\Delta E = 20 \text{ meV}$ obtained from the Arrhenius regime. The result is $I = 100 m_e \text{ nm}^2$, which, assuming a cube of uniform density distributed over the $(\lambda/2)^3$ volume of the fundamental unit, corresponds to 0.1 electron mass (m_e) per chromium unit cell. This remarkable result, derived only from our data and the simple physical picture of Fig. 1, is consistent with Hall effect data^{28,29} showing that the SDW is associated with the loss of a similar number of carriers, which of course must be moved with the rotors when there is a switching event.

We have introduced the direct measurement of noise spectra in antiferromagnets. Our experiments access local mesoscale spin dynamics with just a few domain walls in the illuminated volume, an advantage over non-local experimental probes that cannot be easily applied to macroscopic or bulk structures. The key finding is that even in bulk samples, and at temperatures very low compared to the Néel temperature, domain walls can be visibly unstable on time-scales of fractions of an hour. What this means is that the stability of antiferromagnetism needs to be engineered—for example, by insertion of appropriate pinning centres—into devices that exploit it. This will become even more important for nanoscale spintronics that include antiferromagnetic elements. Beyond the obvious advantages for magnetic engineering of now having a technique with which antiferromagnetic domain fluctuations can be readily assessed, we foresee tremendous opportunities in areas such as the science of antiferromagnetic nanoparticles.

METHODS

Experiments were carried out at beamlines 33-ID and 8-ID of the Advanced Photon Source, Argonne National Laboratory. The undulator-generated X-rays are monochromatized by a Si(111) crystal at an energy $E = 7.35 \text{ keV}$ (wavelength 1.686 \AA). A $10 \mu\text{m}$ pinhole aperture or $10 \mu\text{m}$ (horizontally) by $40 \mu\text{m}$ (vertically) slits placed 5 cm upstream from the sample selected the partially coherent portion of the X-ray beam with a resulting coherence fraction $A \approx 0.07\text{--}0.18$. A high purity Cr(111) wafer (Alfa Aesar) was used to ensure a roughly equal population of domains along the three fundamental cubic axes. The sample was mounted inside a low-drift He flow cryostat, with thermal shielding provided by $600\text{-}\mu\text{m}$ -thick Be dome. Speckle patterns were recorded with a Princeton Instruments PI-LCX 1300 deep depletion X-ray CCD camera ($1,340 \times 1,300$ pixel array, with $20 \mu\text{m} \times 20 \mu\text{m}$ pixel size), located 150 cm from the sample in reflection geometry.

Received 4 December 2006; accepted 20 March 2007.

1. Barkhausen, H. Zwei mit Hilfe der Neuen Verstärker entdeckte Erscheinungen. *Phys. Z.* **20**, 401–403 (1919).
2. Weissman, M. B. Low frequency noise as a tool to study disordered materials. *Annu. Rev. Mater. Sci.* **26**, 395–429 (1996).
3. Sethna, J. P., Dahmen, K. A. & Myers, C. R. Crackling noise. *Nature* **410**, 242–250 (2001).
4. Shull, C. G. & Brockhouse, B. N. in *Nobel Lectures, Physics 1991–1995* (ed. Ekspeng, G.) 107–154 (World Scientific, Singapore, 1997).
5. Vitale, S., Cavalieri, A., Cerdonio, M., Maraner, A. & Prodi, G. A. Thermal equilibrium noise with $1/f$ spectrum in a ferromagnetic alloy: Anomalous temperature dependence. *J. Appl. Phys.* **76**, 6332–6334 (1994).
6. Fawcett, E. Spin-density-wave antiferromagnetism in chromium. *Rev. Mod. Phys.* **60**, 209–283 (1988).
7. Evans, P. G., Isaacs, E. D., Aeppli, G., Cai, Z.-H. & Lai, B. X-ray microdiffraction image of antiferromagnetic domain evolution in chromium. *Science* **295**, 1042–1045 (2002).
8. Michel, R. P., Israeloff, N. E., Weissman, M. B., Dura, J. A. & Flynn, C. P. Electrical-noise measurements on chromium films. *Phys. Rev. B* **44**, 7413–7425 (1991).
9. Sutton, M., Mochrie, S. G. J., Greytak, T., Nagler, S. E. & Berman, L. E. Observation of speckle by diffraction with coherent X-rays. *Nature* **352**, 608–610 (1991).
10. Sutton, M. in *Third-Generation Hard X-Ray Synchrotron Radiation Sources: Source Properties, Optics, and Experimental Techniques* (ed. Mills, D.) 101–123 (Wiley & Sons, New York, 2002).
11. Cipelletti, L., Manley, S., Ball, R. C. & Weitz, D. A. Universal aging features in the restructuring of fractal colloidal gels. *Phys. Rev. Lett.* **84**, 2275–2278 (2000).
12. Bandyopadhyay, R. et al. Evolution of particle-scale dynamics in an aging clay suspension. *Phys. Rev. Lett.* **93**, 228302 (2004).
13. Falus, P., Borthwick, M. A., Narayanan, S., Sandy, A. R. & Mochrie, S. G. J. Cross-over from stretched to compressed exponential relaxations in a polymer-based sponge phase. *Phys. Rev. Lett.* **97**, 066102 (2006).
14. Lemay, S. G., Thorne, R. E., Li, Y. & Brock, J. D. Temporally ordered collective creep and dynamic transition in the charge-density-wave conductor NbSe_3 . *Phys. Rev. Lett.* **83**, 2793–2796 (1999).
15. Fukuyama, H. & Lee, P. A. Dynamics of the charge-density wave. I. Impurity pinning in a single chain. *Phys. Rev. B* **17**, 535–541 (1978).
16. Fukuyama, H. & Lee, P. A. Dynamics of the charge-density wave. II. Long-range Coulomb effects in an array of chains. *Phys. Rev. B* **17**, 542–548 (1978).
17. Littlewood, P. B. & Rice, T. M. Metastability of the \mathbf{Q} vector of pinned charge- and spin-density waves. *Phys. Rev. Lett.* **48**, 44–47 (1982).
18. Cipelletti, L. et al. Universal non-diffusive slow dynamics in aging soft matter. *Faraday Discuss.* **123**, 237–251 (2003).
19. Chudnovsky, E. M. & Tejada, J. *Macroscopic Quantum Tunneling of the Magnetic Moment* (Cambridge University Press, Cambridge, UK, 1998).
20. Barbara, B. et al. Quantum tunneling in magnetic systems of various sizes. *J. Appl. Phys.* **73**, 6703–6706 (1993).
21. Wernsdorfer, W. Classical and quantum magnetization reversal studied in nanometer-sized particles and clusters. *Adv. Chem. Phys.* **118**, 99–190 (2001).
22. Brooke, J., Rosenbaum, T. F. & Aeppli, G. Tunable quantum tunnelling of magnetic domain walls. *Nature* **413**, 610–613 (2001).
23. Fenton, E. W. & Leavens, C. R. The spin density wave in chromium. *J. Phys. F* **10**, 1853–1878 (1980).
24. Fenton, E. W. Domains in the spin-density-wave phases of chromium. *Phys. Rev. Lett.* **45**, 736–739 (1980).
25. Michel, R. P., Weissman, M. B., Ritley, K., Huang, J. C. & Flynn, C. P. Suppression of polarization fluctuations in chromium alloys with commensurate spin-density waves. *Phys. Rev. B* **47**, 3442–3445 (1993).
26. Fullerton, E. E., Bader, S. D. & Robertson, J. L. Spin-density-wave antiferromagnetism of Cr in $\text{Fe/Cr}(001)$ superlattices. *Phys. Rev. Lett.* **77**, 1382–1385 (1996).
27. Hayden, S. M., Doubble, R., Aeppli, G., Perring, T. G. & Fawcett, E. Strongly enhanced magnetic excitations near the quantum critical point of $\text{Cr}_{1-x}\text{V}_x$ and why strong exchange enhancement need not imply heavy fermion behavior. *Phys. Rev. Lett.* **84**, 999–1002 (2000).
28. Lee, M., Husmann, A., Rosenbaum, T. F. & Aeppli, G. High resolution study of magnetic ordering at absolute zero. *Phys. Rev. Lett.* **92**, 187201 (2004).
29. Yeh, A. et al. Quantum phase transition in a common metal. *Nature* **419**, 459–462 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Use of the Center for Nanoscale Materials and Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences. Work at the University of Chicago was supported by the National Science Foundation, while that in London was funded by a Royal Society Wolfson Research Merit Award and the Basic Technologies programme of RCUK.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.G.S. (oshpyrko@anl.gov).

LETTERS

Chemical characterization of element 112

R. Eichler^{1,2}, N. V. Aksenov³, A. V. Belozеров³, G. A. Bozhikov³, V. I. Chepigin³, S. N. Dmitriev³, R. Dressler¹, H. W. Gäggeler^{1,2}, V. A. Gorshkov³, F. Haenssler^{1,2}, M. G. Itkis³, A. Laube¹, V. Ya. Lebedev³, O. N. Malyshev³, Yu. Ts. Oganessian³, O. V. Petrushkin³, D. Piguet¹, P. Rasmussen¹, S. V. Shishkin³, A. V. Shutov³, A. I. Svirikhin³, E. E. Tereshatov³, G. K. Vostokin³, M. Wegrzecki⁴ & A. V. Yeremin³

The heaviest elements to have been chemically characterized are seaborgium¹ (element 106), bohrium² (element 107) and hassium³ (element 108). All three behave according to their respective positions in groups 6, 7 and 8 of the periodic table, which arranges elements according to their outermost electrons and hence their chemical properties. However, the chemical characterization results are not trivial: relativistic effects on the electronic structure of the heaviest elements can strongly influence chemical properties^{4–6}. The next heavy element targeted for chemical characterization is element 112; its closed-shell electronic structure with a filled outer *s* orbital suggests that it may be particularly susceptible to strong deviations from the chemical property trends expected within group 12. Indeed, first experiments concluded that element 112 does not behave like its lighter homologue mercury^{7–9}. However, the production and identification methods^{10,11} used cast doubt on the validity of this result. Here we report a more reliable chemical characterization of element 112, involving the production of two atoms of ²⁸³112 through the alpha decay of the short-lived ²⁸⁷114 (which itself forms in the nuclear fusion reaction¹² of ⁴⁸Ca with ²⁴²Pu) and the adsorption of the two atoms on a gold surface. By directly comparing the adsorption characteristics of ²⁸³112 to that of mercury and the noble gas radon, we find that element 112 is very volatile and, unlike radon, reveals a metallic interaction with the gold surface. These adsorption characteristics establish element 112 as a typical element of group 12, and its successful production unambiguously establishes the approach to the island of stability of superheavy elements through ⁴⁸Ca-induced nuclear fusion reactions with actinides.

Element 112 was discovered¹³ at the Gesellschaft für Schwerionenforschung (GSI), Darmstadt, Germany in 1996. It was produced as an isotope with mass number 277 in the nuclear fusion reaction of ⁷⁰Zn with ²⁰⁸Pb, but its measured half-life of about 200 μ s was too short for a chemical investigation. Three years later, the Flerov Laboratory for Nuclear Reactions (FLNR), Dubna, Russia, announced¹⁴ the production of elements 112 and 114 in nuclear fusion reactions of ⁴⁸Ca with ²³⁸U and ²⁴²Pu. For isotope ²⁸³112, the study indicated¹⁴ a half-life of about 3 min and decay through pure spontaneous fission, confirmed in a later experiment¹⁵. The first studies to chemically characterize this isotope were conducted at FLNR and GSI and suggested^{7–9} that element 112 does not behave like mercury, its lighter homologue in group 12 of the periodic table. However, the identification of element 112 in these studies was uncertain because it relied only on the observation of unspecific spontaneous-fission decay events. A new series of physics experiments at the FLNR in 2004 revealed¹² other decay properties for ²⁸³112: decay through emission of a 9.5-MeV alpha particle and a half-life of about

4 s, followed by the spontaneous fission decay of ²⁷⁹110 with a half-life of about 0.2 s. Moreover, physics experiments at the Lawrence Berkeley National Laboratory, Berkeley, USA¹⁰, and chemistry experiments performed at GSI¹¹ failed to observe element 112 in the nuclear fusion reaction of ⁴⁸Ca with ²³⁸U, which was used in the earlier investigations. In view of these difficulties, for this experiment we used the nuclear fusion of ⁴⁸Ca and ²⁴²Pu, in which ²⁸³112 was reported¹² to be produced as the alpha decay daughter of the primary reaction product ²⁸⁷114 at a rate of several atoms per week.

The systematic order of the periodic table places element 112 in group 12, which also includes zinc, cadmium and mercury. It should thus have the closed-shell electronic ground state configuration Rn: $5f^{14}6d^{10}7s^2$, which implies noble metal characteristics¹⁶. However, relativistic calculations of atomic properties of superheavy elements suggest^{4–6} contraction of the spherical *s*- and *p*_{1/2}-electron orbitals. The effect may increase the chemical stability of the elemental atomic state of element 112 beyond that of a noble metal and endow it with inertness more similar to that of the noble gas radon¹⁷, although recent relativistic calculations on element 112 predicted¹⁸ that it should form a semiconductor-like solid with clear chemical bonds. It was suggested¹⁹ that the questions of the bonding characteristics of element 112 and whether it more strongly resembles a noble metal or a noble gas might be addressed experimentally, by determining its gas adsorption properties on a noble metal surface such as gold.

In fact, relativistic calculations indicate that the spin-orbit splitting of the *6d* orbitals results in element 112 having a ground-state configuration with a *6d*_{5/2} outermost valence orbital, which would make it behave like a noble transition metal^{20,21}. Moreover, relativistic density functional calculations of its interaction with noble metals predict metallic interactions similar to those of the lighter homologue mercury^{22–24}. In the case of large gold clusters and one atom of element 112, the calculations indicate an interaction energy of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{E112}) = 80 \pm 20 \text{ kJ mol}^{-1}$ (refs 23, 24). This result compares well with an adsorption enthalpy estimate from a semi-empirical macroscopic metal–metal adsorption model^{19,25,26}, which yields $-\Delta H_{\text{ads}}^{\text{Au}}(\text{E112}) = 83 \text{ kJ mol}^{-1}$ for a metal-like element 112 on gold^{25,26}; in contrast, the adsorption enthalpy for a noble-gas-like element 112 on gold surfaces was estimated to be $-\Delta H_{\text{ads}}^{\text{Au}}(\text{E112}) = 30 \pm 5 \text{ kJ mol}^{-1}$ (ref. 27). The range of values predicted for the two extreme cases needs to be taken into account when designing the experiment, to ensure that element 112 is detected irrespective of whether it exhibits the adsorption properties of a noble metal or a noble gas.

Thermochromatography allows very efficient probing of the interaction potential of volatile gas-phase species with stationary surfaces over a broad range of interaction enthalpies. We used the *in situ* volatilization and on-line detection method²⁸ for thermochromatography

¹Paul Scherrer Institute, 5232 Villigen, Switzerland. ²Department of Chemistry and Biochemistry, University of Bern, 3012 Bern, Switzerland. ³Flerov Laboratory of Nuclear Reactions, Joint Institute for Nuclear Research, 141980 Dubna, Russia. ⁴Institute of Electron Technology, 02-668 Warsaw, Poland.

measurements at temperatures between $+35\text{ }^{\circ}\text{C}$ and $-186\text{ }^{\circ}\text{C}$, with the original system modified and significantly improved^{11,29} to enable gas adsorption investigations of element 112 on gold surfaces. Figure 1 depicts schematically the experimental set-up. A target of $^{242}\text{PuO}_2$ (1.4 mg cm^{-2} ^{242}Pu) with an admixture of $^{\text{nat}}\text{Nd}_2\text{O}_3$ ($15\text{ }\mu\text{g cm}^{-2}$ of Nd of natural isotopic composition) was deposited on a thin (0.7 mg cm^{-2}) Ti backing foil and irradiated for about three weeks at the U-400 cyclotron at FLNR with 3.1×10^{18} ^{48}Ca particles at a primary energy of $270 \pm 3\text{ MeV}$. The beam energy in the middle of the target was $236 \pm 3\text{ MeV}$, corresponding to the maximum of the production cross-section of $^{287}114$ in the $^{242}\text{Pu}(^{48}\text{Ca}, 3n)$ reaction channel¹². The irradiation generated not only $^{287}114$, but also the partially alpha-decaying nuclide ^{185}Hg with a half-life of 49 s. This nuclide is produced in the reaction $^{142}\text{Nd}(^{48}\text{Ca}, 5n)$ and serves in our experiment as a monitor for the production and separation process. Various isotopes of radon (for example, ^{219}Rn , with a half-life of 4 s) were also produced in multi-nucleon transfer reactions between ^{48}Ca and ^{242}Pu . Thus, radon and mercury were studied simultaneously with element 112 throughout the experiment.

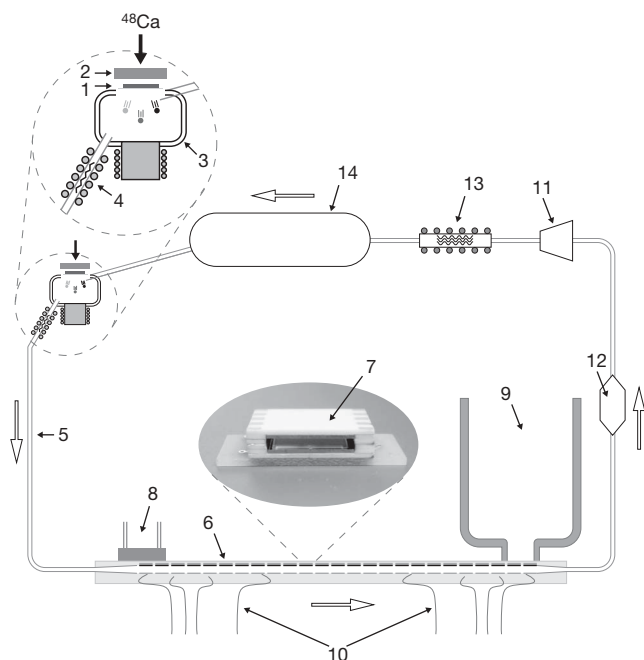


Figure 1 | Schematic experimental set-up used to investigate the adsorption properties of element 112 on a gold surface. A ^{242}Pu target of thickness 1.4 mg cm^{-2} (1) is irradiated by a ^{48}Ca beam that passes through a $4\text{-}\mu\text{m}$ Ti vacuum window (2). The reaction products recoil out of the target through the Ti target backing of $1.5\text{ }\mu\text{m}$ thickness into a recoil stopping volume of 35 cm^3 (3) that is flushed with a 850 ml min^{-1} He/Ar mixture (70/30 vol.%). This carrier gas transported the volatile nuclear reaction products through a quartz wool filter heated to $850\text{ }^{\circ}\text{C}$ (4) and a 8-m-long perfluoroalkoxy-Teflon capillary (5) to the thermochromatographic COLD detector array (6), which consists of 32 pairs of ion-implanted planar silicon detectors facing each other and forming a narrow chromatographic channel (7). One side of the channel is covered with a 50-nm-thick gold layer, deposited directly on the silicon detector surface. A temperature gradient was established along the detector array using a thermostat (8) at the entrance and a liquid nitrogen cryostat (9) at the exit and permanently monitored by eight PT100 platinum resistance thermometers (10) installed along the COLD detector. A self-drying closed gas loop system was developed to keep the amount of trace gases such as oxygen and water in this carrier gas mixture as low as possible. For this purpose the gas flow, enforced by a metal bellows pump (11), is permanently dried by passing it through a drying unit (12) containing Sicapent (a mixture of silica zeolites and phosphorus pentaoxide) and through a getter oven (13) containing tantalum and heated to $1000\text{ }^{\circ}\text{C}$. The gas pulsations originating from the pump are absorbed by a 10-litre buffer volume (14). The white arrows indicate the direction of the gas flow.

The nuclear reaction products passed the backing foil and were stopped in the recoil chamber flushed by the carrier gas. The chamber was covered inside by a quartz inlay (except for the copper beam stop) to avoid losses of element 112 due to adsorption on metallic surfaces. Aerosol particles produced by beam-induced sputtering processes on the copper were stopped directly behind the outlet of the recoil chamber on a quartz wool filter, kept in an oven at $850\text{ }^{\circ}\text{C}$. Additionally, a tantalum foil was inserted into this oven to chemically trap any traces of water and oxygen present in the carrier gas. This set-up guaranteed that only volatile nuclear reaction products entered the perfluoroalkoxy-Teflon capillary (length 8 m, inner diameter 1.56 mm) leading to the detection system. An average transport time from the production place to the detector of about 3.6 s was measured using ^{185}Hg .

The cryo-on-line detector (COLD) consists of an array of 32 pairs of ion-implanted planar silicon detectors ($9.7 \times 9.7\text{ mm}^2$ active area), with the active surfaces facing each other to form a rectangular channel with an open cross-section of $11.6 \times 1.5\text{ mm}^2$. The surface of the detectors on one side of the channel was covered by a 30–50-nm gold layer, while the other side retained the original silicon detector quartz surface that is chemically inert for the elemental state of most elements. The temperature gradient was established along this chromatographic channel by a thermostat heating at the inlet and a liquid-nitrogen cryostat cooling near the outlet. The event-by-event measurement of alpha and spontaneous fission decays provided an on-line identification of nuclides that were deposited on the gold surface in detection geometry close to 4π . Only decays related to the transported volatile nuclides ^{185}Hg , $^{219-221}\text{Rn}$, and $^{209-211}\text{At}$ were observed in the acquired alpha spectra. Mainly owing to the presence of the carrier gas in the detector array the spectroscopic resolution was about 120 keV. The overall transport and separation efficiencies were 90% and 56% for ^{185}Hg ($T_{1/2} = 49\text{ s}$) and ^{219}Rn ($T_{1/2} = 4\text{ s}$), respectively. The detection efficiencies for alpha decay of a species deposited on the detector surface was 87%. Accordingly, the efficiency of detecting an alpha-spontaneous fission (one fragment) or an alpha-spontaneous fission (two coincident fragments) was 87% or 75%, respectively.

During the experiment, two genetically linked decay chains were detected (Fig. 2). They were the only instances for which spontaneous

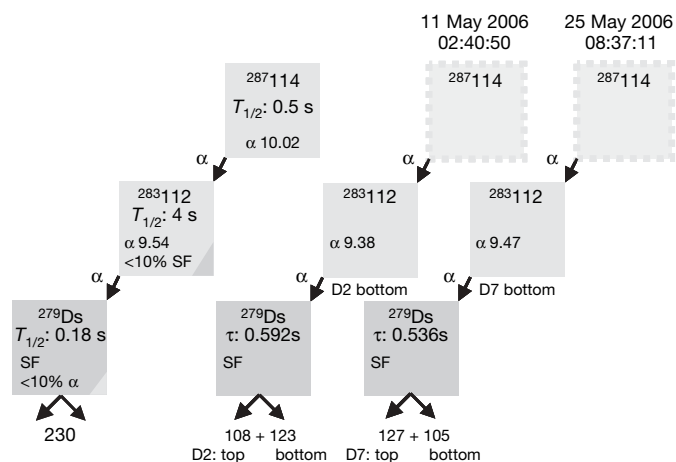


Figure 2 | The observed decay properties of $^{283}112$ and ^{279}Ds . The two decay chains observed in the COLD thermochromatography detector and attributed to the decay of $^{283}112$ are shown on the right side in comparison to the reported decay properties of $^{283}112$ on the left side¹². These decay chains allowed for the unambiguous identification of element 112 after the chemical separation. The alpha- and spontaneous-fission (SF) decay energies are given in MeV. The alpha spectroscopic resolution was about 0.12 MeV. The spontaneous-fission fragment energy resolution is about 20 MeV. The detector (D) number and the distribution of the decay chains in the detector pairs (top or bottom) are indicated. The top detector is gold covered.

fission fragments were measured throughout the entire experiment. The coincident detection of both spontaneous fission fragments revealed a total kinetic energy of about 230 MeV, which is typical for the spontaneous fission decay of a very heavy nucleus¹². Moreover, the observed spontaneous fission events were accompanied by detection of single high-energy alpha decays (9.38 ± 0.12 MeV and 9.47 ± 0.12 MeV, respectively) in the same detector within the preceding 0.6 s; this coincidence implies an exceptionally low probability for the measured decay chains having a random origin (see Supplementary Information section 3). Instead, the observations are consistent with the decay properties of $^{283}\text{112}$ and ^{279}Ds inferred¹² from the experiments with the Dubna gas-filled recoil separator (Fig. 2). The current experiment did not confirm the initial observation of $^{283}\text{112}$ and its decay purely by spontaneous fission with a half-life of about 3 min^{14,15}.

The primary fusion product $^{287}\text{114}$ has a half-life of about 0.5 s, making it too short-lived to be transported to the COLD detector and thus allowing only the decay of the daughter nuclei $^{283}\text{112}$ and ^{279}Ds to be detected. Still, this observation represents a first independent confirmation of the production of element 112 and Ds in the nuclear fusion reaction of ^{48}Ca with ^{242}Pu . Furthermore, the independent confirmation of the decay properties of the isotope $^{283}\text{112}$ and its daughter ^{279}Ds substantiates the discovery in the more recent FLNR experiments¹² that $^{287}\text{114}$ and $^{291}\text{116}$ decay via alpha particle emission to $^{283}\text{112}$ and ^{279}Ds , respectively.

The observation that the two atoms of element 112 were transported through the experimental set-up to their deposition positions is indicative of a high volatility of this element (see also Supplementary Information section 1). In the first part of the experiment the inlet of the detector array was held at -24°C , allowing the cold end of the detector to reach its minimum temperature of -184°C and thus ensuring efficient deposition of the major part of radon-like species in the detector. Figure 3a gives the distribution of ^{185}Hg , ^{219}Rn and $^{283}\text{112}$ deposited in the detector array. About

88% of the ^{219}Rn produced in the experiment was deposited on the last eight detectors, while the other 12% decayed in-flight when passing through detectors 1 to 24 (that is, about 0.5% decayed per detector pair). Deposition of ^{185}Hg on the gold surface occurred in the first detectors, that is, between -24°C and -50°C (Fig. 3a). This pattern is indicative of spontaneous deposition controlled only by the diffusion of ^{185}Hg out of the carrier gas to the gold surfaces. Under these conditions the first atom of $^{283}\text{112}$ was detected on the second detector at a temperature of -28°C (Fig. 3a, black arrow).

The observed deposition pattern in COLD is reproduced in kinetic Monte Carlo simulations of the gas-adsorption chromatography experiment³⁰, making it possible to infer the adsorption enthalpies ($-\Delta H_{\text{ads}}^{\text{Au}}$) for the elements on gold (see also Supplementary Information, section 1). In the case of mercury, the observed deposition pattern on the gold surface gives a lower limit of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Hg}) > 47 \text{ kJ mol}^{-1}$, while successful transport of the element at room temperature (20°C) through 8 m perfluoroalkoxy-Teflon capillary, in conjunction with empirical correlations, provides an upper limit of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Hg}) < 125 \text{ kJ mol}^{-1}$ (see Supplementary Information, section 1). Both limits are in agreement with literature data of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Hg}) = 98 \pm 3 \text{ kJ mol}^{-1}$ (ref. 29). The deposition pattern of radon translates into an adsorption enthalpy of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Rn}) = 20 \pm 1 \text{ kJ mol}^{-1}$, which is somewhat lower than the literature value²⁷ of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Rn}) = 29 \pm 3 \text{ kJ mol}^{-1}$ for radon adsorption on gold surfaces. But it agrees well with the measured value of $-\Delta H_{\text{ads}}^{\text{ice}}(\text{Rn}) = 20 \pm 2 \text{ kJ mol}^{-1}$ for radon adsorption on ice surfaces³¹. This agreement and the observation that the dew point in the carrier gas was colder than -95°C points to a water content of $< 0.1 \text{ p.p.m.}$, suggesting that a thin ice layer covers the surfaces of the detectors that are kept at temperatures below -95°C . The observation in the first part of the experiment of element 112 in the main deposition region of mercury suggests a high probability for this atom having been deposited upon first contact with the gold surface.

The second part of the experiment was performed after raising the temperature at the inlet of the detector array to 35°C , the maximum temperature allowing safe operation of the silicon detectors. The increased inlet temperature leads to a temperature of -180°C at the cold end of the thermochromatography channel (Fig. 3b). Under these conditions about 65% of the ^{219}Rn deposited on the last five detectors, which according to the kinetic Monte Carlo model corresponds to a value of $-\Delta H_{\text{ads}}^{\text{ice}}(\text{Rn}) = 20 \pm 1 \text{ kJ mol}^{-1}$ that is consistent with the first measurement. For mercury, the diffusion-controlled deposition pattern of ^{185}Hg on the gold surface within the first eight detectors corresponds to a lower limit of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Hg}) > 65 \text{ kJ mol}^{-1}$, again consistent with the known adsorption data for mercury on gold²⁹. The second atom of $^{283}\text{112}$ to be observed in this study deposited on the seventh detector, held at -5°C . We note that only about 5% of the entire mercury reached this detector or deposited at lower temperatures (Fig. 3b).

The alpha particle decay of the $^{283}\text{112}$ atoms was detected both times on the non-gold-covered side of the detector pairs, which then also detected the subsequent spontaneous fission decay attributed to ^{279}Ds . This suggests that the alpha-decay daughter ^{279}Ds was recoil-implanted into the gold-covered detector owing to the momentum transfer during the alpha particle emission from $^{283}\text{112}$; otherwise, the recoiling ^{279}Ds would have been stopped in the carrier gas between the detectors and transported downstream to another detector pair. Therefore, we conclude that $^{283}\text{112}$ was deposited on the gold-covered side of the detector (see also Supplementary Information section 1).

The statistical Monte Carlo approach to modelling the gas chromatography results^{17,30} uses adsorption enthalpy values to mimic the observed deposition patterns, which provides upper and lower limits for the adsorption enthalpy of element 112 on gold $-\Delta H_{\text{ads}}^{\text{Au}}(\text{E112})$ of 98 kJ mol^{-1} and 45 kJ mol^{-1} (68% confidence interval), respectively; it also yields a most probable value of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{E112}) = 52 \text{ kJ mol}^{-1}$, which has a large associated uncertainty due to the

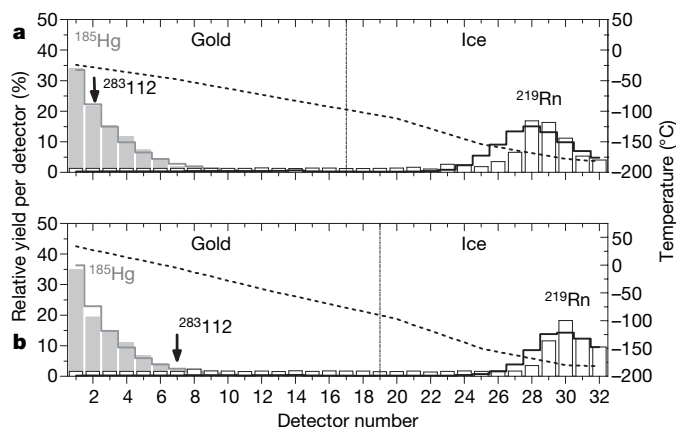


Figure 3 | Results of thermochromatographic separations of element 112 compared to mercury and radon in the COLD detector. The relative yields (left-handed axis) of ^{185}Hg (grey bars) and ^{219}Rn (white bars) are shown together with the two observed decays of $^{283}\text{112}$ (black arrows) as a function of the detector number. The results of a microscopic model of the adsorption process based on a Monte Carlo approach³⁰ are presented (solid lines). The vertical dashed-dotted lines indicate the beginning of the ice coverage of the gold surface at temperatures below -95°C . **a**, First part of the experiment: Monte Carlo simulations with adsorption enthalpies of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Hg}) > 47 \text{ kJ mol}^{-1}$ (grey solid line) and $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Rn}) = 20 \text{ kJ mol}^{-1}$ (black solid line), are shown, respectively. The black dashed line (right-handed axis) represents the temperature gradient applied (-24 to -184°C). **b**, Second part of the experiment: Monte Carlo simulations with adsorption enthalpies of $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Hg}) > 65 \text{ kJ mol}^{-1}$ (grey solid line) and $-\Delta H_{\text{ads}}^{\text{Au}}(\text{Rn}) = 20 \text{ kJ mol}^{-1}$ (black solid line) are shown, respectively. The black dashed line (right-handed axis) represents the temperature gradient applied ($+35$ to -180°C).

small number of observed events (see also Supplementary Information section 2). Still, the range of likely adsorption enthalpies inferred from this study indicates an interaction between element 112 and gold that is significantly stronger than the purely dispersive van der Waals interactions of noble-gas like elements²⁷. We therefore conclude that the stronger adsorption interaction of element 112 with gold involves formation of a metal bond, which is behaviour typical of group 12 elements.

Received 27 July 2006; accepted 19 March 2007.

- Schädel, M. *et al.* Chemical properties of element 106 (seaborgium). *Nature* **388**, 55–57 (1997).
- Eichler, R. *et al.* Chemical characterization of bohrium (element 107). *Nature* **407**, 63–65 (2000).
- Düllmann, Ch. E. *et al.* Chemical investigation of hassium (Hs, element 108). *Nature* **418**, 859–862 (2002).
- Fricke, B. Superheavy elements. *Structure Bonding* **21**, 90–144 (1975).
- Pyykkö, P. & Desclaux, J.-P. Relativity and the periodic system of elements. *Acc. Chem. Res.* **12**, 276–281 (1979).
- Schwerdtfeger, P. & Seth, M. in *Encyclopaedia of Computational Chemistry* Vol. 4 2480–2499 (Wiley, New York, 1998).
- Yakushev, A. B. *et al.* First attempt to chemically identify element 112. *Radiochim. Acta* **89**, 743–745 (2001).
- Yakushev, A. B. *et al.* Chemical identification and properties of element 112. *Radiochim. Acta* **91**, 433–440 (2003).
- Gäggeler, H. W. *et al.* Chemical and nuclear studies of hassium and element 112. *Nucl. Phys. A* **734**, 208–212 (2004).
- Gregorich, K. E. *et al.* Attempt to confirm superheavy element production in the ⁴⁸Ca + ²³⁸U reaction. *Phys. Rev. C* **72**, 014605 (2005).
- Eichler, R. *et al.* Attempts to chemically investigate element 112. *Radiochim. Acta* **94**, 181–191 (2006).
- Oganessian, Yu. Ts. *et al.* Measurements of cross sections and decay properties of the isotopes of elements 112, 114, and 116 produced in the fusion reactions ^{233,238}U, ²⁴²Pu, and ²⁴⁸Cm + ⁴⁸Ca. *Phys. Rev. C* **70**, 064609 (2004).
- Hofmann, S. *et al.* The new element 112. *Z. Phys. A* **354**, 229–230 (1996).
- Oganessian, Yu. Ts. *et al.* Synthesis of nuclei of the superheavy element 114 in reactions induced by ⁴⁸Ca. *Nature* **400**, 242–245 (1999).
- Oganessian, Yu. Ts. *et al.* Second experiment at VASSILISSA separator on the synthesis of the element 112. *Eur. Phys. J. A* **19**, 3–6 (2004).
- Eichler, B. Volatility properties of transactinides around Z=114 (prediction). *Kernenergie* [in German] **19**, 307–311 (1976).
- Pitzer, K. S. Are elements 112, 114, and 118 relatively inert gases? *J. Chem. Phys.* **63**, 1032–1333 (1975).
- Gaston, N., Opahle, I., Gäggeler, H. W. & Schwerdtfeger, P. Is eka-mercury (element 112) a group 12 metal? *Angew. Chem. Int. Edn* **46** (14) 2444–2447 (2007).
- Eichler, B. & Rossbach, H. Adsorption of volatile metals on metal surfaces and its application in nuclear chemistry. I. Calculation of adsorption enthalpies for hypothetical superheavy elements with Z around 114. *Radiochim. Acta* **33**, 121–125 (1983).
- Eliav, E., Kaldor, U. & Ishikawa, Y. Transition energies of mercury and ekamercury (element 112) by the relativistic coupled-cluster method. *Phys. Rev. A* **52**, 2765–2769 (1995).
- Seth, M., Schwerdtfeger, P. & Dolg, M. The chemistry of the superheavy elements. I Pseudopotentials for 111 and 112 and relativistic coupled clusters calculations for (112)H⁺, (112)F₂, and (112)F₄. *J. Chem. Phys.* **106**, 3623–3632 (1997).
- Pershina, V., Bastug, T., Jacob, T., Fricke, B. & Varga, S. Intermetallic compounds of the heaviest elements: the electronic structure and bonding of dimers of element 112 and its homolog Hg. *Chem. Phys. Lett.* **365**, 176–183 (2002).
- Sarpe-Tudoran, C. *et al.* Adsorption of super-heavy elements on metal surfaces. *Eur. Phys. J. D* **24**, 65–67 (2003).
- Pershina, V. & Bastug, T. Relativistic effects on experimentally studied gas-phase properties of the heaviest elements. *Chem. Phys.* **311**, 139–150 (2005).
- Eichler, B. *Metal Chemistry of Transactinides*. PSI Report 00–09, ISSN 1019–0643 (Villigen, 2000).
- Eichler, B. *Volatilization Properties of Transactinides from Metal Surfaces and Melts (Thermochemical Calculation)*. PSI Report 03–01, ISSN 1019–0643 (Villigen, 2002).
- Eichler, R. & Schädel, M. Adsorption of radon on metal surfaces: a model study for chemical investigations of elements 112 and 114. *J. Phys. Chem. B* **106**, 5413–5420 (2002).
- Düllmann, Ch. E. *et al.* IVO, a device for in situ volatilization and on-line detection of products from heavy ion reactions. *Nucl. Instrum. Meth. A* **479**, 631–639 (2002).
- Soverna, S. *et al.* Thermochemical studies of mercury and radon on transition metal surfaces. *Radiochim. Acta* **93**, 1–8 (2005).
- Zvara, I. Simulation of thermochromatographic processes by the Monte Carlo method. *Radiochim. Acta* **38**, 95–101 (1985).
- Eichler, B., Zimmermann, P. & Gäggeler, H. W. Adsorption of radon on ice surfaces. *J. Phys. Chem. A* **104**, 3126–3131 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to the staff of the U-400 cyclotron for providing intense beams of ⁴⁸Ca. This work was supported in part by the Russian Foundation for Basic Research and by the Swiss National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.E. (robert.eichler@psi.ch).

LETTERS

A scaling law for slow earthquakes

Satoshi Ide¹, Gregory C. Beroza², David R. Shelly² & Takahiko Uchide¹

Recently, a series of unusual earthquake phenomena have been discovered, including deep episodic tremor¹, low-frequency earthquakes², very-low-frequency earthquakes³, slow slip events⁴ and silent earthquakes^{5–9}. Each of these has been demonstrated to arise from shear slip, just as do regular earthquakes, but with longer characteristic durations and radiating much less seismic energy. Here we show that these slow events follow a simple, unified scaling relationship that clearly differentiates their behaviour from that of regular earthquakes. We find that their seismic moment is proportional to the characteristic duration and their moment rate function is constant, with a spectral high-frequency decay of f^{-1} . This scaling and spectral behaviour demonstrates that they can be thought of as different manifestations of the same phenomena and that they comprise a new earthquake category. The observed scale dependence of rupture velocity for these events can be explained by either a constant low-stress drop model or a diffusional constant-slip model. This new scaling law unifies a diverse class of slow seismic events and may lead to a better understanding of the plate subduction process and large earthquake generation.

In recent years, an expanding variety of unusual earthquakes have been discovered. Since the 1990s, silent earthquakes^{5–9}, some with moment magnitude M_w as large as 7.5, have been discovered through Global Positioning System and strain meter measurements, with durations ranging from minutes to months. In particular, a dense network of high-sensitivity seismometers and tiltmeters maintained by the National Research Institute for Earth Science and Disaster Prevention, Hi-net¹⁰, has successively led to the discovery of low-frequency earthquakes (LFE)², low-frequency tremor¹, and slow-slip events (SSE)⁴ in essentially the same place along the Nankai trough in western Japan. Similar phenomena have been found in the Cascadia subduction zone¹¹. These episodes have come to be referred to as episodic tremor and slip (ETS). A different type of event, known as very-low-frequency earthquakes (VLF), with characteristics between those of LFEs and SSEs, has now also been discovered in western Japan³. Despite the variety of names, these events occur at the same time and in the same place, which suggests a close relationship and perhaps a common origin. A unifying characteristic of these events is that they have much longer durations than ordinary earthquakes of comparable seismic moment. For this reason, we refer to them all as slow earthquakes.

Figure 1 shows examples of these events around Shikoku, Japan. Both LFEs and low-frequency tremor occur in a belt-like zone that follows the 35-km iso-depth curves of the top of the Philippine Sea plate¹². Although LFEs are isolated in time and tremor is continuous, these two phenomena radiate seismic waves with very similar waveforms and spectra, predominantly at frequencies of 1–8 Hz, suggesting that the underlying physical mechanism is the same¹³. LFEs appear to be shear slip on the plate interface because of their distribution on an $\sim 20^\circ$ dipping plane that corresponds to the dip of Philippine Sea plate subduction¹⁴ and because they have a sense of

deformation consistent with relative plate motion¹⁵. SSEs are detected by tiltmeters and co-located with LFEs and tremor. Two types of SSEs have been reported, long-term SSEs of $M_w \approx 7$ and short-term SSEs of $M_w \approx 6$. A long-term SSE in 2003 continued for about five months, while short-term SSEs occur over 3–7-day periods⁴. VLFs are weak signals found in both borehole accelerometer (tiltmeter) records and in broadband seismograms. Their locations overlap those of the other phenomena, but with a characteristic frequency of 0.02–0.05 Hz (ref. 3). The fault slip models of SSEs and the moment tensors of VLFs are consistent with the mechanisms of LFEs, past megathrust earthquakes¹⁶, and global plate motion¹⁷. Therefore it appears that this entire range of phenomena represents shear slip on the plate interface.

Table 1 and Fig. 2 summarize characteristic quantities of low-frequency tremor, LFEs, VLFs and SSEs in the Nankai trough, together with those of ETS in the Cascadia subduction zone and some silent earthquakes. Except for two silent earthquakes that occurred along the Japan trench^{6,7} in places where tremor activity is not reported, we find a clear proportionality between the characteristic duration T (in units of seconds) and the seismic moment, a measure of the size of final fault motion, M_0 (in units of N m):

$$M_0 \approx T \times 10^{12-13} \quad (1)$$

which is different from the scaling of regular (in the sense of ‘normal’) earthquakes in subduction zones¹⁸:

$$M_0 \approx T^3 \times 10^{15-16} \quad (2)$$

Since the discovery of silent earthquakes, the possibility of a gap between large, regular earthquakes and silent earthquakes has become apparent. We propose that this gap is there because the two types of earthquake represent different modes of slip propagation, and that the only events capable of filling the gap would be a composite of two rupture modes.

If we assume that the proportionality between seismic moment and characteristic time holds at all times during an earthquake, then the seismic moment rate function $\dot{M}_0(t)$ would be constant until the event’s termination. Approximately constant moment rates have been observed at large scales, both in the Nankai subduction zone⁸ and in numerical simulations¹⁹. Far-field seismic displacement $u(t)$ is proportional to $\dot{M}_0(t)$, so we would expect a box-car displacement pulse for these events. The Fourier amplitude spectrum of a box-car function of duration T is proportional to $\sin(\pi f T)/(\pi f T)$, where f is frequency. This spectrum is asymptotically flat at low-frequency and falls asymptotically as f^{-1} at high-frequency. Although the actual moment rate function in detail will be more complex than this, a seismic source with fairly abrupt onset and termination compared to variations in moment rate should lead to a spectrum that is similar.

This spectral behaviour differs from that observed for regular earthquakes, which exhibit f^{-2} spectral decay at high frequency

¹Department of Earth and Planetary Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan. ²Department of Geophysics, 397 Panama Mall, Stanford University, Stanford, California 94305-2215, USA.

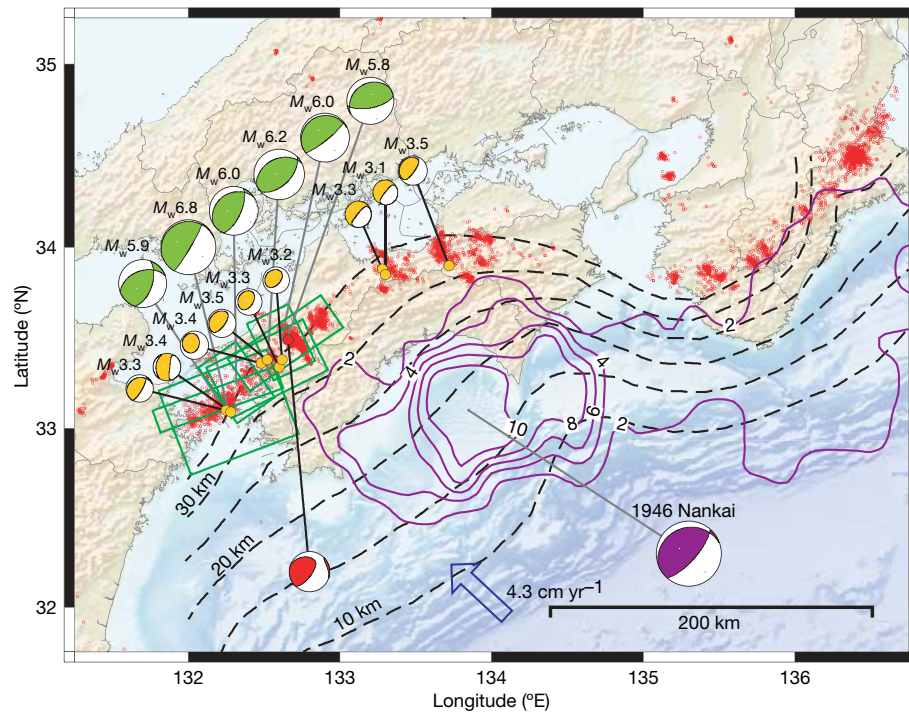


Figure 1 | Various types of earthquakes and their mechanisms along the Nankai trough, western Japan. Red dots represent LFE locations determined by the Japan Meteorological Agency². Red and orange 'beach balls' show the mechanism of LFEs¹⁵ and VLFs³, respectively. Green rectangles and beach balls show fault slip models of SSEs⁴. Purple contours

and the purple beach ball show the slip distribution³¹ (in metres) and focal mechanism¹⁶ of the 1946 Nankai earthquake ($M_w = 8$). The top of the Philippine Sea plate is shown by dashed contours¹². The blue arrow represents the direction of relative plate motion in this area¹⁷.

(Fig. 3a). The spectra of VLFs, LFEs and tremor all follow f^{-1} decay at high frequency (Fig. 3b, c), which suggests a constant moment rate over these smaller length scales too. The f^{-1} decay is also supported by the difference in slope between stacked spectra of LFEs and regular earthquakes¹³. That is, the spectra of these events are consistent with the overall proportionality between released seismic moment and elapsed time during the slow earthquakes shown in Fig. 2. Figure 3b also suggests that the observed characteristic size of VLFs and LFEs may be due to the strong background noise from microseisms, which inhibits the discovery of events with sizes intermediate between VLFs and LFEs. If tremor, LFEs, VLFs and SSEs are all governed by the same process in the same area, the difference between them corresponds only to the final size that an event achieves. Our scaling relations suggest that we should expect slower earthquakes of $M_w \approx 4-5$ with durations of tens of minutes to an hour in the same area. Such events have not yet been observed, but may be detectable on sensitive broadband instruments, if noise levels are sufficiently low.

The discovery of this scaling relation leads to the question of what controls these processes. Seismic moment can be written using the characteristic dimension of the fault plane L and the average slip amount D :

$$M_0 = \mu DL^2 \quad (3)$$

where μ is the rigidity. The slip amounts of SSEs and ETSs are of the order of centimetres. Because the average slip amounts of smaller events are unknown, we consider two mechanisms with different scalings of D . First, we assume that D is proportional to L as in regular earthquakes. Stress drop $\Delta\sigma \approx \mu D/L$ is about 10 kPa for the large slow earthquakes in Table 1 if μ is 30 GPa, so this assumption leads to a constant stress drop of 10 kPa. This is smaller by about two orders of magnitude than that of regular earthquakes and we refer to this assumption as the low-stress drop model. The constant-stress drop assumption, together with equations (1) and (3), gives the relation

Table 1 | Source parameters of slow earthquakes

Type	M_0 (N m)	T (s)	Characteristic frequency (Hz)	D (m)	L (m)	Area
Low-frequency tremor			1–8			Nankai trough
LFE	3×10^{11}	0.3	1–8			Nankai trough ¹⁵
VLF	$0.4-2.2 \times 10^{14}$	~ 20	0.02–0.05			Nankai trough ³
SSE short-term	$0.6-2.8 \times 10^{18}$	$3-6 \times 10^5$		0.008–0.026	$3-5 \times 10^4$	Nankai trough ⁴
SSE long-term	1.7×10^{19}	10^7		0.11	$5-7 \times 10^4$	Nankai trough ⁴
Silent earthquakes	1.1×10^{19}	3×10^7		< 0.18	6×10^4	Nankai trough ⁵
	1.4×10^{19}	5×10^7		0.2	10^5	Nankai trough ⁸
	$1-4 \times 10^{20}$	10^5				Northeast Japan ⁶
	3×10^{20}	10^6				Northeast Japan ⁷
	2.2×10^{20}	1.7×10^7		0.1	$2-5 \times 10^5$	Mexico ⁹
ETS	$0.7-2.0 \times 10^{19}$	$2-3 \times 10^6$		0.02	1.5×10^5	Cascadia ¹¹
Afterslip	4.2×10^{20}	3×10^7		0.7–0.9	10^5	Northeast Japan ²⁷
Shallow VLF	$0.4-1.6 \times 10^{15}$	~ 10	0.08–0.24			Northeast Japan ²⁶
Slow slip (creep)	2×10^{16}	2×10^5		0.03–0.1	5,000	San Andreas ²²
	4.4×10^{13}	3,600				San Andreas ²¹
Slow slip in volcano	6.8×10^{17}	1.9×10^5		> 0.015	1.5×10^4	Hawaii ²⁵
Aseismic fault	4.9×10^{13}	100			250	Italy ^{23,24}

between L and T :

$$L^3 = C_1 T \quad (4)$$

where $C_1 \approx 3 \times 10^8 \text{ m}^3 \text{ s}^{-1}$. This relation predicts that the characteristic propagation velocity for these events L/T is proportional to $1/L^2$; it is fast for shorter lengths and slow for longer lengths, which is consistent with the observed propagation velocity of these slow earthquakes. When $L \approx 100 \text{ km}$, the characteristic velocity is about $5\text{--}15 \text{ km day}^{-1}$ ($\sim 0.1 \text{ m s}^{-1}$) for ETS and SSE, while tremor propagates for about 10 km at a rupture velocity of about 45 km h^{-1} ($\sim 10 \text{ m s}^{-1}$)¹³. Although the above assumption describes the observed size-dependent rupture velocity, the physical mechanism that leads to this dependence is unclear.

As an alternative model, we next assume that the slip amounts are almost constant, and limited by the accumulated plate motion since the last event. That is, we assume they are comparable to the value of slip observed in SSEs and ETSs. Assuming a constant D for all events, equations (1) and (3) lead to a relation between L and T :

$$L^2 = C_2 T \quad (4)$$

where the constant C_2 is about $10^4 \text{ m}^2 \text{ s}^{-1}$ if $D = 0.01 \text{ m}$ and μ is 30 GPa . Because this is of the form of a scaling relation for diffusion-controlled physical phenomena²⁰ with a diffusion constant of C_2 , we refer to this relationship as the diffusional earthquake model. In this model, the stress drop $\Delta\sigma \approx \mu D/L$ is larger for an initial small event and decreases with increasing event size. The propagation velocity L/T decreases proportionally to $1/L$, as the stress concentration decreases. Equation (4) explains the rupture velocity of 0.1 m s^{-1}

for $L \approx 100 \text{ km}$, while the characteristic velocity of 1 m s^{-1} for $L \approx 10 \text{ km}$ is somewhat slower than observed.

It is interesting to note that the scaling relations of slow and regular earthquakes intersect at about $M_w \approx 1$ ($M_0 \approx 10^{11} \text{ N m}$), near the minimum size of both LFEs and typically detected regular earthquakes. In the low-stress earthquake model, the rupture propagation velocity would be $\sim 10 \text{ km s}^{-1}$, which is unphysical because it would be faster than the elastic wave speed. The diffusional earthquake model predicts a rupture velocity of 1 km s^{-1} with the stress drop of about 20 MPa , which is not impossible, but in the case of stress drop is much higher than observed for regular earthquakes. These physically unrealizable conditions may limit the smallest scale of the slow earthquakes to the magnitude of LFEs or the slightly smaller tremor. The lack of VLFs or SSEs during some LFE and tremor episodes suggests that a single LFE or slightly smaller events may be an elementary event that marks the smallest slow earthquake.

As more and more earthquakes that fit into the category of slow earthquakes are identified, it is natural that they would initially be considered as different and perhaps unrelated phenomena. In this regard, we note that there are other earthquakes falling into this category that occur in other environments. Transient slip events have

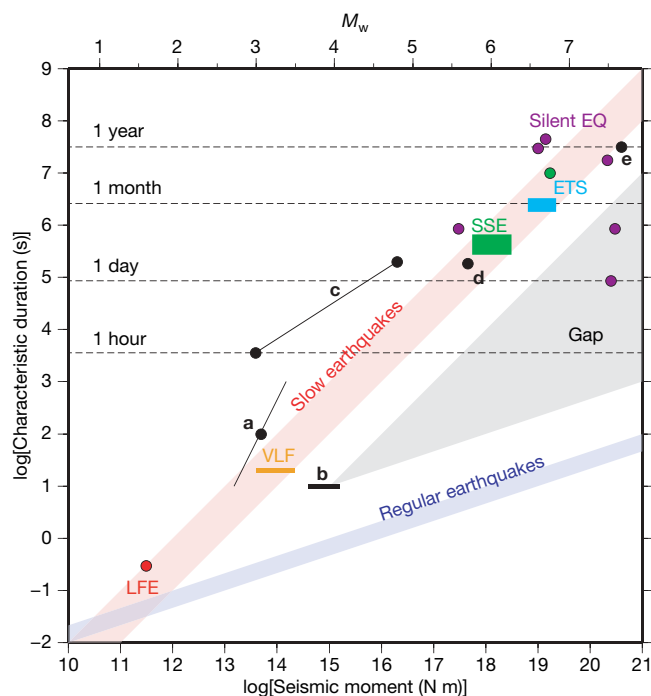


Figure 2 | Comparison between seismic moment and the characteristic duration of various slow earthquakes in Table 1. LFE (red), VLF (orange), and SSE (green) occur in the Nankai trough while ETS (light blue) occur in the Cascadia subduction zone. These follow a scaling relation of $M_0 \propto t$, for slow earthquakes. Purple circles are silent earthquakes. Black symbols are slow events listed in the bottom half of Table 1. **a**, Slow slip in Italy^{23,24}, representing a typical event (circle) and proposed scaling (line). **b**, VLF earthquakes in the accretionary prism of the Nankai trough²⁶. **c**, Slow slip and creep in the San Andreas Fault^{21,22}. **d**, Slow slip beneath Kilauea volcano²³. **e**, Afterslip of the 1992 Sanriku earthquake²⁷. Typical scaling relation for shallow interplate earthquakes is also shown by a thick blue line.

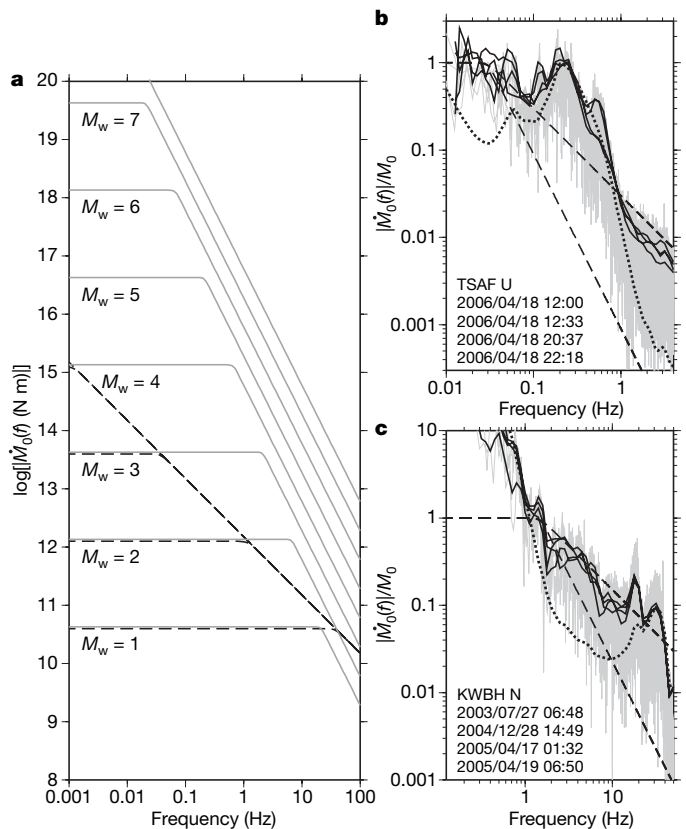


Figure 3 | Characteristics of moment rate spectrum of slow earthquakes. **a**, Scaling of moment rate spectrum of slow earthquakes (black dashed lines) and regular earthquakes (grey lines) for different magnitudes. **b**, An example of displacement spectrum (moment rate spectrum) of a VLF event. Original (grey) and averaged (thick black) spectra are shown. The seismograms represent the up-down (U) component at station TSAF for four VLFs that occurred on 18 April 2006 at the times shown in Table S1 of ref. 3. The noise level is shown by the dotted line. There is a large peak due to microseisms between $0.1\text{--}1 \text{ Hz}$. **c**, Same as **b** but for an LFE. The seismograms represent the north-south (N) component at station KWBH for the reference LFE in ref. 15 and three LFEs that occurred nearby, at the times shown. (Note that the frequency axes in **b** and **c** are different.) For both VLF and LFE, spectra are well explained with a slope of f^{-1} at high frequency rather than f^{-2} (dashed curves). We corrected for the effect of attenuation, assuming that these waves are S waves and $Q = 200$. Peaks above 10 Hz probably show amplification due to the site and/or instrument characteristics.

been discovered along the San Andreas Fault, California^{21,22}, along an inactive fault in Italy^{23,24}, beneath Kilauea volcano, Hawaii²⁵, and within the shallow accretionary prism along the Japan trench²⁶. The afterslip of large earthquakes, such as that observed after the 1994 Sanriku earthquake²⁷, may also be regarded as a kind of silent slip, and thus fall into this category. The duration and seismic moment of these events all fall in or around the diffusional limb of the scaling law shown in Fig. 2. The substantial scatter in this log-arithmetic plot indicates that other physical phenomena may well be important, but the difference in scaling between regular and slow earthquakes defines two distinct populations, and is likely to be a useful starting point for classifying different phenomena.

Important questions remain about the physical mechanism behind both the low-stress drop and the diffusional earthquake models. In the slow-earthquake zone of the Nankai trough, the existence of fluid is suggested by seismic wave tomography¹⁴, which suggests that the introduction of fluid and its subsequent diffusion is the underlying mechanism. Rate- and state-dependent friction laws^{28,29} can also lead to propagation of slow slip without fluids^{19,30}. Further studies will address these questions and undoubtedly refine the simplifications made in our interpretation; however, the scaling law for slow earthquakes unifies a seemingly diverse class of unusual, recently discovered seismic events. Understanding these earthquakes should lead to new insights into the physics of plate subduction and in assessing the probability of future large earthquakes.

Received 8 December 2006; accepted 26 March 2007.

- Obara, K. Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* **296**, 1679–1681 (2002).
- Katsumata, A. & Kamaya, N. Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL015981 (2003).
- Ito, Y., Obara, K., Shiomi, K., Sekine, S. & Hirose, H. Slow earthquakes coincident with episodic tremors and slow slip events. *Science* **26**, 503–506 (2006).
- Hirose, H. & Obara, K. Repeating short- and long-term slow slip events with deep tremor activity, around the Bungo channel region, southwest Japan. *Earth Planets Space* **57**, 961–972 (2005).
- Hirose, H., Hirahara, K., Kimata, F., Fujii, N. & Miyazaki, S. A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, southwest Japan. *Geophys. Res. Lett.* **26**, 3237–3240 (1999).
- Kawasaki, I. et al. The 1992 Sanriku-oki, Japan, ultra-slow earthquake. *J. Phys. Earth* **43**, 105–116 (1995).
- Kawasaki, I., Asai, Y. & Tamura, Y. Space-time distribution of interplate moment release including slow earthquakes and seismo-geodetic coupling in the Sanriku-oki region along the Japan trench. *Tectonophysics* **330**, 267–283 (2001).
- Ozawa, S. et al. Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan. *Science* **298**, 1009–1012 (2002).
- Kostoglodov, V. et al. A large silent earthquake in the Guerrero seismic gap, Mexico. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017219 (2003).
- Obara, K., Kasahara, K., Hori, S. & Okada, Y. A densely distributed high-sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. *Rev. Sci. Instrum.* **76**, doi:10.1063/1.1854197 (2005).
- Dragert, H., Wang, K. & Rogers, G. Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth Planets Space* **56**, 1143–1150 (2004).
- Baba, T., Tanioka, T., Cummins, P. R. & Uehira, K. The slip distribution of the 1946 Nankai earthquake estimated from tsunami inversion using a new plate model. *Phys. Earth Planet. Inter.* **132**, 59–73 (2002).
- Shelly, D. R., Beroza, G. C. & Ide, S. Non-volcanic tremor and low frequency earthquake swarms. *Nature* **446**, 305–307 (2007).
- Shelly, D. R., Beroza, G. C., Ide, S. & Nakamura, S. Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature* **442**, 188–191 (2006).
- Ide, S., Shelly, D. R. & Beroza, G. C. The mechanism of deep low frequency earthquakes: further evidence that deep non-volcanic tremor is generated by shear slip on the plate interface. *Geophys. Res. Lett.* **34**, doi:10.1029/2006GL028890 (2007).
- Ando, M. A fault model of the 1946 Nankaido earthquake derived from tsunami data. *Phys. Earth Planet. Inter.* **28**, 320–336 (1982).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. *Geophys. Res. Lett.* **21**, 2191–2194 (1994).
- Houston, H. Influence of depth, focal mechanism, and tectonic setting on the shape and duration of earthquake source time functions. *J. Geophys. Res.* **106**, 11137–11150 (2001).
- Shibazaki, B. & Iio, Y. On the physical mechanism of silent slip events along the deeper part of the seismic zone. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017047 (2003).
- Barenblatt, G. I. *Scaling, Self-similarity, and Intermediate Asymptotics* (Cambridge Univ. Press, Cambridge, UK, 1996).
- Gladwin, M. T., Gwyther, R. L., Hart, R. H. G. & Brechenridge, K. S. Measurements of the strain field associated with episodic creep events on the San Andreas fault at San Juan Bautista, California. *J. Geophys. Res.* **99**, 4559–4564 (1994).
- Linde, A. T., Gladwin, M. T., Johnston, M. J. S., Gwyther, R. L. & Bilham, R. G. A slow earthquake sequence on the San Andreas fault. *Nature* **383**, 65–67 (1996).
- Crescentini, L., Amoruso, A. & Scarpa, R. Constraints on slow earthquake dynamics from a swarm in central Italy. *Science* **286**, 2132–2134 (1999).
- Amoruso, A., Crescentini, L., Morelli, A. & Scarpa, R. Slow rupture of an aseismic fault in a seismogenic region of Central Italy. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL016027 (2002).
- Segall, P., Desmarais, E. K., Shelly, D., Miklius, A. & Cervelli, P. Earthquakes triggered by silent slip events on Kilauea volcano, Hawaii. *Nature* **442**, 71–74 (2006).
- Ito, Y. & Obara, K. Very low frequency earthquakes within accretionary prisms are very low-stress-drop earthquakes. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL025883 (2006).
- Heki, K., Miyazaki, S. & Tsuji, H. Silent fault slip following an interplate thrust earthquake at the Japan Trench. *Nature* **386**, 595–597 (1997).
- Dieterich, J. H. Modelling of rock friction: I Experimental results and constitutive equations. *J. Geophys. Res.* **84**, 2161–2168 (1979).
- Ruina, A. L. Slip instability and state variable friction laws. *J. Geophys. Res.* **88**, 10,359–10,370 (1983).
- Liu, Y. & Rice, J. R. Aseismic slip transients emerge spontaneously in three-dimensional rate and state modeling of subduction earthquake sequences. *J. Geophys. Res.* **110**, doi:10.1029/2004JB003424 (2005).
- Sagiya, T. & Thatcher, W. Coseismic slip resolution along a plate boundary megathrust: the Nankai Trough, southwest Japan. *J. Geophys. Res.* **104**, 1113–1129 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Vidale for comments. This work is supported by a Grant-in-Aid for Scientific Research, the Ministry of Education, Sports, Science and Technology, Japan, and the National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.I. (ide@eps.s.u-tokyo.ac.jp).

LETTERS

Drought sensitivity shapes species distribution patterns in tropical forests

Bettina M. J. Engelbrecht^{1,2}, Liza S. Comita³, Richard Condit¹, Thomas A. Kursar^{1,4}, Melvin T. Tyree^{1,5,6}, Benjamin L. Turner¹ & Stephen P. Hubbell^{1,3}

Although patterns of tree species distributions along environmental gradients have been amply documented in tropical forests^{1–7}, mechanisms causing these patterns are seldom known. Efforts to evaluate proposed mechanisms have been hampered by a lack of comparative data on species' reactions to relevant axes of environmental variation¹. Here we show that differential drought sensitivity shapes plant distributions in tropical forests at both regional and local scales. Our analyses are based on experimental field assessments of drought sensitivity of 48 species of trees and shrubs, and on their local and regional distributions within a network of 122 inventory sites spanning a rainfall gradient across the Isthmus of Panama. Our results suggest that niche differentiation with respect to soil water availability is a direct determinant of both local- and regional-scale distributions of tropical trees. Changes in soil moisture availability caused by global climate change and forest fragmentation are therefore likely to alter tropical species distributions, community composition and diversity.

Ecologists have long recognized correlations between water availability and species distributions⁸. Indeed, the most consistent patterns of species distributions in tropical forests worldwide, as well as in many non-tropical systems, are regional and continental correlations with rainfall^{1,4–7,9}, and local associations with topography, which frequently drives variation in soil water availability^{2,3,10}. Most tropical forests experience regular dry seasons once or twice per year, and even in many aseasonal equatorial areas extended dry spells occur, during which plants may suffer drought stress¹¹. Variation in water availability—especially during dry periods—mediated through rainfall, soil characteristics or topography, is therefore an important candidate for niche partitioning in tropical forest species. However, other factors that may co-vary with water availability have also been proposed to lead to the observed patterns, namely light and nutrient availability, or herbivore and pathogen pressure^{1,7,12}. Furthermore, species distributions can result from limited dispersal coupled with speciation, delayed responses to climate change, or other historical effects¹³. To advance our understanding beyond pattern description and determine which mechanisms are important, we need experimental assessments of the responses of species to potentially limiting factors, and a linking of species-specific responses to species distributions at different scales¹.

We have previously shown experimentally that seedlings of tropical woody plants in central Panama vary considerably in their drought sensitivity¹⁴. Other studies have documented nonrandom species distributions with respect to water availability at both regional and local scales across the Isthmus of Panama^{3,4}. Here we set out to test the hypothesis that variation in drought sensitivity explains these distribution patterns. We link our extensive comparative data set on

seedling drought sensitivity to quantitative data on the regional distribution of species across the rainfall gradient in central Panama, and to their local distribution with respect to wet and dry habitats in a 50-ha permanent forest plot. We further test whether correlations

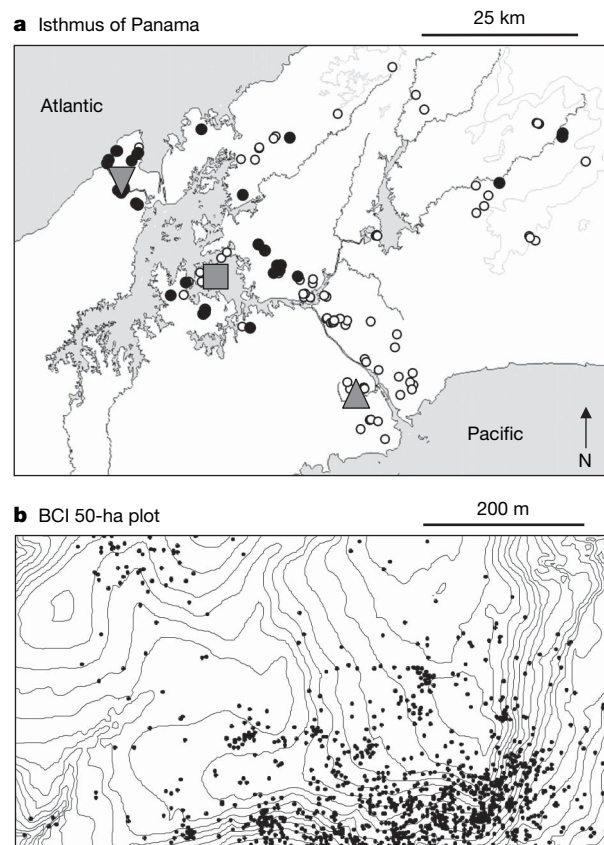


Figure 1 | Maps of the study areas and distribution of one tree species, as an example. **a**, Location of the study plots along the Panama Canal. The Cocoli plot on the dry Pacific side (upright, filled triangle), the Sherman plot at the wet Atlantic side (inverted, filled triangle), the 50-ha forest dynamics plot on BCI (filled square), and the 119 additional inventory sites (circles) are shown. *Xylopia macrantha* (Annonaceae), a relatively drought-sensitive species (Supplementary Information), occurs on the wet Caribbean, but not on the drier Pacific, side of the isthmus (filled circles, present; open circles, absent). **b**, Topographic map of the 50-ha forest dynamics plot (contour interval 2 m). *Xylopia macrantha* (black circles) shows a higher density on wet slopes than in dry plateau habitats.

¹Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancon, Panama. ²Department of Plant Ecology and Systematics, University of Kaiserslautern, 67663 Kaiserslautern, Germany. ³Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA. ⁴Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA. ⁵USDA Forest Service, PO Box 968, Burlington, Vermont 05402, USA. ⁶Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2E3, Canada.

between drought sensitivity and species distributions might arise indirectly as a result of differences in the shade tolerance of species due to a trade-off between drought and shade tolerance¹⁵.

To assess drought sensitivity we conducted irrigation experiments on transplanted seedlings of 48 native tree and shrub species. The difference in survival between dry and irrigated conditions ranged from 0% to 100% (see Supplementary Information), confirming that species vary widely and continuously in drought sensitivity. In separate studies we showed that variation in seedling drought sensitivity is due to species differences in the effectiveness of physiological mechanisms of drought tolerance (that is, their differential ability to survive despite low leaf water potentials¹⁶) (T.A.K. *et al.*, unpublished data).

To assess regional species distributions we first examined the densities of species in two large plots (4–5 ha) on opposite sides of the Isthmus of Panama (Fig. 1a). We found that species' density at the dry Pacific side relative to the wet Atlantic side correlated negatively with experimentally assessed drought sensitivity (Fig. 2a, $r^2 = 0.44$, $n = 23$, $P = 0.0006$). Analysis of an independent data set of species occurrence in a network of 122 sites across the isthmus (Fig. 1a) also showed that experimental drought sensitivity was a significant predictor of the probability of occurrence of the species on the dry relative to the wet side (Fig. 2c, $r^2 = 0.19$, $n = 39$, $P = 0.006$); species exhibiting high sensitivity to drought in the field experiments occurred more often towards the wet end of the climatic gradient.

At the local scale we assessed species' densities on dry plateau versus wet slope sites within the Barro Colorado Island (BCI) 50-ha plot (Fig. 1b). Both for seedlings and trees, densities on the dry plateau relative to wet slopes were significantly and negatively correlated with drought sensitivity (Fig. 2b, d; seedlings: $r^2 = 0.14$,

$n = 33$, $P = 0.035$; adults: $r^2 = 0.34$, $n = 33$, $P = 0.0004$). Drought sensitivity explained more than twice the variation in density in dry versus wet sites for adult trees than it did for seedlings (34% versus 14%, Fig. 2b, d). In addition, more species showed associations with the wetter slope habitat as adults than as juveniles in the BCI plot¹⁷. This implies that, at local scales, seeds arrive and seedlings are able to establish in sites with suboptimal soil moisture conditions, and differential mortality across species and habitats subsequently enhances associations with favourable habitats.

Although tropical plant species' reactions to environmental factors, namely light and nutrients, have been experimentally assessed in numerous studies, only a few have quantitatively linked these data to distribution patterns^{1,18,19}. These studies were restricted to a small number of species, precluding analysis of the importance of environmental factors across the community. Thus, the findings presented here represent the most thorough study so far linking tropical plant species distribution patterns with species' reactions to an environmental factor at the community level.

Our results suggest that drought sensitivity has a direct role in determining species distributions with respect to local and regional water availability. However, significant correlations between drought sensitivity and distributions might arise indirectly as a result of differences in the shade tolerance of species if there is a trade-off between shade and drought tolerance (as proposed in ref. 15) and if forest light availability decreases as water availability increases.

We found no correlation between light requirements and species' densities (or probability of occurrence) in dry versus wet habitats at the local or regional scale (see Supplementary Information). In addition, we found no correlation between the drought sensitivity of species and their light requirements (Supplementary Information). This is in contrast to the results of a recent meta-analysis²⁰ that found a trade-off between drought tolerance and shade tolerance for temperate species. On the basis of our data, we can rule out the possibility that the observed correlations between tropical species distributions and drought sensitivity indirectly result from differential species' light requirements. This is despite the obvious importance that species' responses to light have for the distribution of species in young successional versus mature forests and in canopy gap versus understorey sites.

Observed correlations between species distributions and drought sensitivity might also be driven by correlations between water and nutrient availability, coupled with a correlation between species drought sensitivity and nutrient requirements. Sufficient experimental data on the nutrient requirements of species in this system are lacking. However, preliminary data on soil nutrients and pH suggest that water and nutrient availability are not significantly correlated (see Supplementary Information). This does not rule out the possibility that nutrient availability is involved in regulating the distributions of species in tropical forests²¹, but indicates that it is unlikely that the significant correlations between species distribution patterns and drought sensitivity are driven by variation in nutrient availability.

The results presented here emphasize the sensitivity of tropical forests to water availability²² and suggest that future changes in hydrological processes and precipitation patterns will have direct consequences for species ranges, tropical forest community composition and ecosystem function. Such changes in rainfall patterns have been observed in parts of the tropics over the past decades and are projected to continue and intensify as a consequence of forest fragmentation and global climate change²³. Current vegetation–climate models, particularly for tropical regions, suffer from a lack of ecological data and mechanistic understanding of the factors shaping current species distributions. The knowledge that water availability, together with the sensitivity of species to drought, is a causal mechanistic factor influencing species distribution patterns in tropical forests will help to improve the accuracy and specificity of predictions of vegetation shifts under global change scenarios.

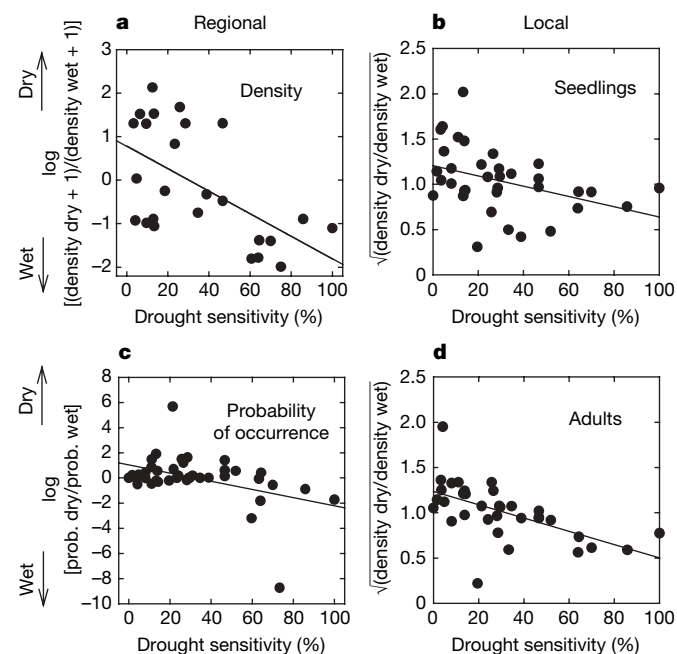


Figure 2 | Significant relationships between drought sensitivity of seedlings and regional and local distributions of species. a, c, Regional distributions of tree and shrub species were assessed as the density in a plot at the dry Pacific relative to the wet Atlantic side of the Isthmus of Panama (a), and as the probability of occurrence on the dry side relative to the wet side of the isthmus based on data from 122 inventory sites (c). (Note that results were similar when omitting the two outlying species in c.) b, d, Local habitat affinities with topographic units within the 50-ha forest dynamics plot on BCI were assessed for seedlings (b) and adults (d) as the density of a species in dry plateau sites relative to wet slope and streamside habitats. Drought sensitivity was experimentally assessed as the relative survival difference between dry and irrigated conditions (see Supplementary Information).

METHODS SUMMARY

We assessed regional species distributions across the Isthmus of Panama (Fig. 1a), which spans a strong gradient of annual rainfall²⁴ and dry season length (Supplementary Data 1 and 2), in two ways. First, we calculated the density of adults of each species at a dry site on the Pacific side of the isthmus relative to a wet site located on the Atlantic side²⁵. Second, we analysed presence/absence data from 122 inventory sites across the isthmus^{4,5,25,26} (Supplementary Data 3). An index of dry season response for 44 of the experimental species was based on the fitted probability of occurrence towards the dry end (dry season of 145 days) and towards the wet end of the gradient (110 days; Supplementary Data 4 and Supplementary Program 1). Local-scale species distributions were assessed within a 50-ha forest dynamics plot on BCI²⁷ at the centre of the rainfall gradient (Fig. 1b). We analysed local associations with dry and wet habitats as species density on dry plateaux relative to wet slopes for seedlings²⁸ and adults²⁷ (Supplementary Information).

Species' drought sensitivity for seedlings of 48 native tree and shrub species was experimentally assessed in the field as the relative difference in survival over the dry season of transplanted seedlings in dry versus irrigated plots^{14,29} (see Supplementary Information). As an index of the light requirements of species, we used observational data on the percentage of recruits encountered in low canopy sites (that is, high light conditions in gaps) in the BCI 50-ha plot³⁰. Soil pH and the concentrations of nitrogen and phosphorous were measured at 19 of the inventory sites across the rainfall gradient with standard procedures (see Supplementary Information).

Linear regressions were used to examine whether species' reactions to drought or light were significant predictors of their densities (or probability of occurrence) in dry versus wet sites.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 November 2006; accepted 14 March 2007.

- Veenendaal, E. M. & Swaine, M. D. Limits to tree species distribution in lowland tropical rainforests. In *Dynamics of Tropical Communities* (eds Newbery, D. M., Prins, H. H. T. & Brown, N.) 163–191, 37th Symp. Brit. Ecol. Soc. (Blackwell, Oxford, 1998).
- Webb, C. O. & Peart, D. R. Habitat associations of trees and seedlings in a Bornean rain forest. *J. Ecol.* **88**, 464–478 (2000).
- Harms, K. E., Condit, R., Hubbell, S. P. & Foster, R. B. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *J. Ecol.* **89**, 947–959 (2001).
- Pyke, C. R., Condit, R., Aguilar, S. & Lao, S. Floristic composition across a climatic gradient in a neotropical lowland forest. *J. Veg. Sci.* **12**, 553–566 (2001).
- Condit, R. *et al.* Beta-diversity in tropical forest trees. *Science* **295**, 666–669 (2002).
- Bongers, F., Poorter, L., Van Rompaey, R. S. A. R. & Parren, M. P. E. Distribution of twelve moist forest canopy tree species in Liberia and Cote d'Ivoire: response curves to a climatic gradient. *J. Veg. Sci.* **10**, 371–382 (1999).
- ter Steege, H. *et al.* A spatial model of tree α -diversity and tree density for the Amazon. *Biodiv. Conserv.* **12**, 2255–2277 (2003).
- Schimper, A. F. W. *Plant Geography upon a Physiological Basis* (Clarendon, Oxford, 1903).
- Gentry, A. H. Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Mo. Bot. Gard.* **75**, 1–34 (1988).
- Daws, M. I., Mullins, C. E., Burslem, D. F. R. P., Paton, S. R. & Dalling, J. W. Topographic position affects the water regime in a semideciduous tropical forest in Panama. *Plant Soil* **238**, 79–90 (2002).
- Walsh, R. P. D. Climate. In *The Tropical Rain Forest an Ecological Study* 2nd edn (ed. Richards, P. W.) 159–202 (Cambridge Univ. Press, Cambridge, USA, 1996).
- Givnish, T. J. On the causes of gradients in tropical tree diversity. *J. Ecol.* **87**, 193–210 (1999).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).

- Engelbrecht, B. M. J. & Kursar, T. A. Comparative drought-resistance of seedlings of 28 species of co-occurring tropical woody plants. *Oecologia* **136**, 383–393 (2003).
- Huston, M. A. *Biological Diversity. The Coexistence of Species on Changing Landscapes* (Cambridge Univ. Press, Cambridge, UK, 1994).
- Tyree, M. T., Engelbrecht, B. M. J., Vargas, G. & Kursar, T. A. Desiccation tolerance of five tropical seedlings in Panama: Relationship to a field assessment of drought-performance. *Plant Physiol.* **132**, 1439–1447 (2003).
- Comita, L. S., Condit, R. & Hubbell, S. P. Developmental changes in habitat associations of tropical trees. *J. Ecol.* **95**, 482–492 (2007).
- Veenendaal, E. M. *et al.* Responses of West African forest tree seedlings to irradiance and soil fertility. *Funct. Ecol.* **10**, 501–511 (1996).
- Palmiotto, P. A. *et al.* Soil-related habitat specialization in Dipterocarp rain forest tree species in Borneo. *J. Ecol.* **92**, 609–623 (2004).
- Ninemets, Ü. & Valladares, F. Tolerance to shade, drought, and waterlogging of temperate northern hemisphere trees and shrubs. *Ecol. Monogr.* **76**, 521–547 (2006).
- John, R. *et al.* Soil nutrients influence spatial distributions of tropical tree species. *Proc. Natl Acad. Sci. USA* **104**, 864–869 (2007).
- Potts, M. D. Drought in a Bornean everwet rain forest. *J. Ecol.* **91**, 467–474 (2003).
- Malhi, Y. & Phillips, O. L. Tropical forests and global atmospheric change: a synthesis. *Phil. Trans. R. Soc. Lond. B* **359**, 549–555 (2004).
- Smithsonian Tropical Research Institute. ACP physical monitoring downloads. (http://striweb.si.edu/esp/physical_monitoring/download_acp.htm) (2006).
- Condit, R. *et al.* Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. *J. Trop. Ecol.* **20**, 51–72 (2004).
- Pérez, R. *et al.* Tree species composition and diversity in the Upper Chagres River Basin, Panama. In *The Rio Chagres, Panama: A Multidisciplinary Profile of a Tropical Watershed* (ed. Harmon, R. S.) (Springer, Dordrecht, The Netherlands, 2005).
- Condit, R., Hubbell, S. P. & Foster, R. B. Barro Colorado Forest census plot data. (<http://ctfs.si.edu/datasets/bci/>) (2005).
- Comita, L. S., Aguilar, S., Pérez, R., Lao, S. & Hubbell, S. P. Patterns of woody plant species abundance and diversity in the seedling layer of a tropical forest. *J. Veg. Sci.* **18**, 163–174 (2007).
- Engelbrecht, B. M. J., Kursar, T. A. & Tyree, M. T. Drought effects on seedling survival in a tropical moist forest. *Trees (Berl.)* **19**, 312–321 (2005).
- Welden, C. W., Hewett, S. W., Hubbell, S. P. & Foster, R. B. Sapling survival, growth, and recruitment: Relationship to canopy height in a neotropical forest. *Ecology* **72**, 35–50 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the more than two dozen field assistants who conducted the extensive censuses and drought experiments. S. Aguilar and R. Perez were instrumental in species identifications, and S. Lao assisted with data management. The study was financially supported by the Andrew W. Mellon Foundation (T.A.K., B.M.J.E. and M.T.T.), the National Science Foundation (S.P.H. and L.S.C.), the University of Utah (T.A.K. and B.M.J.E.), the Center of Tropical Forest Science (B.M.J.E. and L.S.C.) and the German Science Foundation (B.M.J.E.). We acknowledge logistical support provided by the Smithsonian Tropical Research Center and the opportunity to work in the area granted by the 'Autoridad Nacional del Ambiente' (ANAM) of Panama.

Author Contributions B.M.J.E. and L.S.C. analysed the compiled data sets and coordinated the writing. B.M.J.E., T.A.K. and M.T.T. conducted the drought sensitivity experiments. L.S.C. and S.P.H. managed the seedling census, and R.C. and S.P.H. the tree census, in the 50-ha forest dynamics plot. R.C. coordinated the establishment and census of the regional inventory network, and modelled species occurrence against dry season length. B.L.T. conducted soil nutrient analyses. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.M.J.E. (engelbrb@si.edu).

METHODS

Regional species distributions. We assessed regional species distributions across the Isthmus of Panama, which spans a strong rainfall gradient from ~1,600 to 4,800 mm yr⁻¹ over a distance of 65 km (ref. 22). Annual rainfall is highly correlated with dry season length ($r^2 = 0.90$, $n = 44$, $P < 0.0001$; Supplementary Information) and minimum dry season gravimetric soil water content ($r^2 = 0.60$, $n = 19$, $P = 0.0002$; B.M.J.E., unpublished data). Regional distributions of species were assessed two ways. First, we examined the density of each species (≥ 1 cm diameter at breast height, d.b.h.) at a dry site on the Pacific side of the isthmus (Cocoli, 4 ha, Fig. 1a) and at a wet site located on the Atlantic side (Fort Sherman, 4.96 ha), using data collected in 1998²⁵. We calculated the ratio of species' density in the Pacific-side plot to the Atlantic-side plot for all species with ≥ 20 total individuals. Second, we analysed presence/absence data from 122 inventory sites across the isthmus^{4,5,25,26} (Fig. 1a and Supplementary Information) for species occurring in ≥ 10 of the sites. We quantified dry season duration based on data from rainfall stations near the Panama Canal²⁴ and used this to model dry season length at the sites (Supplementary Information). The responses of 44 of the experimental tree species were modelled as a function of dry season duration, fitting a kernel using an optimized bandwidth³¹. This approach requires no a priori assumptions about the shape of the response, and produces an estimated probability of occurrence for each species at any dry season duration. An index of dry season response for each species was based on the fitted probability of occurrence towards the dry end (dry season of 145 days) and towards the wet end of the gradient (110 days). For details see Supplementary Information.

Local species distributions. Local-scale species distributions were assessed within a 50-ha forest dynamics plot on Barro Colorado Island (BCI)²⁶ at the centre of the rainfall gradient (Fig. 1b). At BCI, as in other tropical forests, dry season soil moisture is lower on plateaux than slopes^{10,32}. We used data on species' densities from a 2001 census of stems ≥ 20 cm height to 1 cm d.b.h. (referred to as seedlings)²⁸ and from a 2000 census of stems ≥ 1 cm d.b.h. (referred to as trees)²⁷. We analysed local associations with dry and wet habitats as species density on plateaux ('high' and 'low' plateaux combined, as defined by ref. 3) relative to slopes (slopes and streamside areas combined) for all species with ≥ 10 individuals within the focal habitats for both life stages (Supplementary Information).

Seedling drought sensitivity. The drought sensitivity of species was experimentally assessed in the Barro Colorado Nature Monument, in the centre of the isthmus, for seedlings of 48 native tree and shrub species from 29 families (expanded from ref. 14; Supplementary Information). We transplanted first-year seedlings to 60 plots in the forest understorey of which half were kept dry with transparent rainout-shelters and the other half were irrigated (see ref. 14 for experimental details). As an index of species' drought sensitivity (DS) (that is, their susceptibility to periods of low water availability) we calculated the relative difference in survival (S) over the dry season in dry versus irrigated plots²⁹ ($DS = (S_{\text{irrigated}} - S_{\text{dry}}) / S_{\text{irrigated}} \times 100$; Supplementary Information). The experimental approach allowed for the differentiation of effects of low water availability from other factors potentially inducing mortality during the dry season, such as shade, herbivore damage or nutrient limitation²⁹. Experiments were conducted in the dry seasons of 2000/2001 and 2002/2003, with eight species overlapping between experiments. After 11 weeks in 2002/2003, relative differential survival of the overlapping species correlated highly significantly ($r^2 = 0.89$, $n = 8$, $P = 0.0004$) and closest to 1:1 with relative differential survival after 22 weeks in 2000/2001 (regression not different from 1:1, t -test $P > 0.05$). Values of drought sensitivity measured in 2003 (DS_{2003}) were adjusted using the equation of the regression between relative survival in the two years for the eight overlapping species ($DS_{\text{adjusted}} = 6.492 + 0.903 \times DS_{2003}$), yielding a comparable data set for all 48 species.

Plant light requirements. As an index of the light requirements of species we used observational data on the percentage of recruits encountered in low canopy sites (that is, high light conditions in gaps) in the BCI 50-ha plot³⁰. A high proportion of recruits in low canopy sites suggests that the species is light-demanding. In contrast to the experimental data on drought sensitivity, this data set does not allow us to differentiate among possible factors contributing to differences in recruitment rates between gaps and shaded understorey sites (for example, light requirements, pest resistance, etc.).

Soil pH and nutrient content. Soil pH and the concentrations of nitrogen and phosphorus were analysed in 19 of the 1-ha plots across the isthmus for the surface 10 cm of soil sampled at 25 systematically distributed sites per plot. The cores from each plot were combined before analysis. Soil pH was measured in a 1:2 ratio of soil to deionized water. Total nitrogen was determined by combustion and gas chromatography using a Flash 1112 Soil Analyser (CE Elantech). Total phosphorus was determined by ignition and acid digestion, with detection by inductively coupled plasma optical emission spectrometry. Total nitrogen

and phosphorus concentrations were correlated tightly with measures of 'plant-available' nitrogen and phosphorus, respectively (B.L.T. and B.M.J.E., unpublished data).

Linking data sets. Linear regressions were used to examine whether species' drought sensitivity or light requirement was a significant predictor of their densities (or probability of occurrence) in dry versus wet sites. Sample sizes depended on the number of species overlapping between respective data sets. Data were transformed to meet requirements of normal distribution and constant variance (see figures).

31. McCune, B. Non-parametric habitat models with automatic interactions. *J. Veg. Sci.* **17**, 819–830 (2006).
32. Becker, P., Rabenold, P. E., Idol, J. R. & Smith, A. P. Water potential gradients for gaps and slopes in a Panamanian tropical moist forest's dry season. *J. Trop. Ecol.* **4**, 173–184 (1988).

Intrinsic functional architecture in the anaesthetized monkey brain

J. L. Vincent^{1,5,6}, G. H. Patel^{1,2,3}, M. D. Fox¹, A. Z. Snyder^{1,2}, J. T. Baker³, D. C. Van Essen³, J. M. Zempel², L. H. Snyder³, M. Corbetta^{1,2,3} & M. E. Raichle^{1,2,3,4}

The traditional approach to studying brain function is to measure physiological responses to controlled sensory, motor and cognitive paradigms. However, most of the brain's energy consumption is devoted to ongoing metabolic activity not clearly associated with any particular stimulus or behaviour¹. Functional magnetic resonance imaging studies in humans aimed at understanding this ongoing activity have shown that spontaneous fluctuations of the blood-oxygen-level-dependent signal occur continuously in the resting state. In humans, these fluctuations are temporally coherent within widely distributed cortical systems that recapitulate the functional architecture of responses evoked by experimentally administered tasks^{2–6}. Here, we show that the same phenomenon is present in anaesthetized monkeys even at anaesthetic levels known to induce profound loss of consciousness. We specifically demonstrate coherent spontaneous fluctuations within three well known systems (oculomotor, somatomotor and visual) and the 'default' system, a set of brain regions thought by some to support uniquely human capabilities. Our results indicate that coherent system fluctuations probably reflect an evolutionarily conserved aspect of brain functional organization that transcends levels of consciousness.

It has been suggested that coherent spontaneous blood-oxygen-level-dependent (BOLD) fluctuations observed in the resting state reflect unconstrained but consciously directed mental activity⁷. Alternatively, coherent BOLD fluctuations may persist in the absence of normal perception and behaviour, reflecting a more fundamental or intrinsic property of functional brain organization. Importantly, the former view predicts that coherent BOLD fluctuations should be absent in the anaesthetized state, in which conscious mental activity is thought to be absent.

We acquired whole-brain BOLD functional magnetic resonance imaging (fMRI) in isoflurane-anaesthetized monkeys and observed highly organized patterns of spontaneous BOLD fluctuations (see Methods). We first examined ongoing fluctuations in the oculomotor system because it is well characterized both functionally and anatomically in the monkey and shows correlated spontaneous BOLD fluctuations in awake resting humans⁸. To explore the spatial distribution of ongoing BOLD fluctuations within the monkey oculomotor system, we correlated the time course of BOLD fluctuations averaged within a priori defined regions of interest (ROIs) around the frontal eye fields (FEF) and the lateral intraparietal area (LIP) with the time course of each voxel within the brain (regions defined from a study of visually guided saccades⁹). Figure 1 shows the distribution of voxels temporally correlated with fluctuations in the right LIP. Correlation maps were computed on the basis of single BOLD runs (15 min each) in one monkey at two levels of isoflurane

anaesthesia (A, 0.90%; B, 1.25%) and are shown paired with the simultaneously acquired electroencephalogram (EEG). These traces show continuous slow waves of EEG activity (Fig. 1a) as well as burst-suppression patterns (Fig. 1b), which are characteristic of moderately deep anaesthesia. The principal constituents of the cortical oculomotor system (bilateral FEF and bilateral LIP) are present in the correlation maps at both levels of anaesthesia. These single-subject results demonstrate that coherent spontaneous BOLD fluctuations persist during the anaesthetized state. All subsequently presented results represent data acquired over multiple BOLD runs in multiple animals at isoflurane levels between 0.8% and 1.5%. Group-averaged ($N = 8$) statistical maps for each oculomotor seed ROI are shown in Supplementary Fig. 1. Conjunction analyses of these data demonstrated that FEF, LIP, middle temporal (MT) and middle superior temporal (MST) areas were consistently correlated with all oculomotor seed ROIs (Fig. 2a and Supplementary Fig. 1e).

To determine the system specificity of coherent spontaneous BOLD fluctuations, additional seed ROIs were anatomically defined in left and right somatomotor cortex (SMC; see Supplementary Fig. 2). Regional time courses of BOLD fluctuations were extracted from all

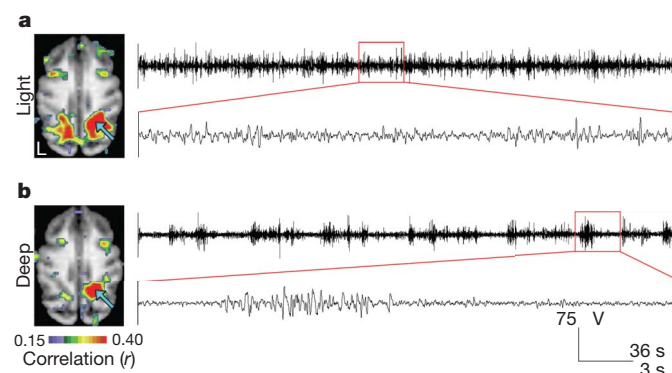


Figure 1 | BOLD correlation maps obtained at two levels of isoflurane anaesthesia in one monkey. Correlation maps computed with a seed region in right LIP (blue arrows) are superimposed on a structural MRI. Simultaneously recorded EEG is shown to the right of each correlation map. In both **a** and **b**, the top and bottom traces show 360 and 30 s of EEG, respectively. **a**, 0.90% isoflurane; continuous, slow EEG activity. L, left side. **b**, 1.25% isoflurane; burst suppression EEG pattern. The principal components of the cortical oculomotor system (bilateral FEF, bilateral LIP) are present in the correlation maps during both light (**a**) and deep (**b**) anaesthesia (defined electrophysiologically). A more systematic examination of the dependence of BOLD correlations on level of anaesthesia is given in Supplementary Figs 4–6.

¹Departments of Radiology, ²Neurology, ³Anatomy and Neurobiology, and ⁴Biomedical Engineering, Washington University in St Louis, Missouri 63110, USA. ⁵Department of Psychology, Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. ⁶Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA.

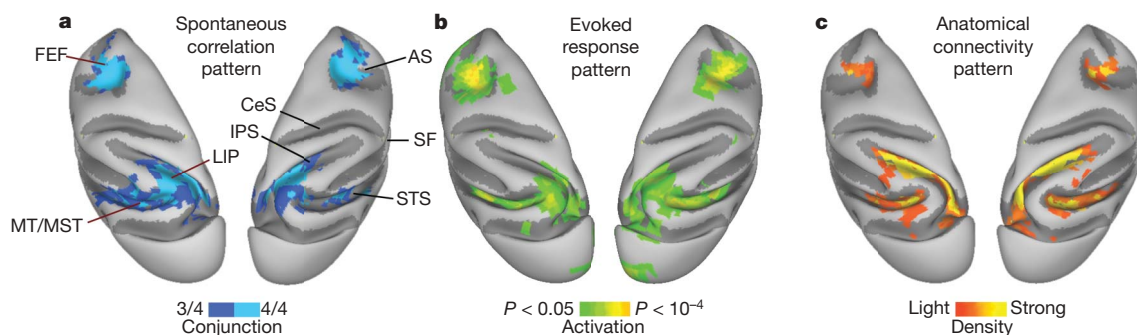


Figure 2 | Cortical patterns of coherent spontaneous BOLD fluctuations are similar to those of task-evoked responses and anatomical connectivity. **a**, Conjunction map of BOLD correlations within the oculomotor system on dorsal views of the monkey atlas left and right hemisphere surfaces (same data as Supplementary Fig. 1e). Voxels significantly correlated with three (dark blue) or four (light blue) oculomotor ROIs are shown. **b**, Activation pattern evoked by performance of a saccadic eye movement task (average of two monkeys; adapted from ref. 9). **c**, Density of cells labelled by retrograde

tracer injections into right LIP (average of three monkeys; adapted from ref. 10). The left hemisphere injection data are duplicated by reflection of the right hemisphere to facilitate visual comparison. Only regions that showed reproducible projections to LIP (three out of three monkeys) are shown. AS, arcuate sulcus; CeS, central sulcus; IPS, intraparietal sulcus; SF, sylvian fissure; STS, superior temporal sulcus. Data are available at <http://sumsdb.wustl.edu/sums/directory.do?id=6602646>.

oculomotor and somatomotor ROIs and cross-correlated (Supplementary Fig. 3a). Spontaneous BOLD fluctuations were robustly temporally correlated between each pair of oculomotor ROIs (two-tailed t -test, all $P < 0.05$) and similarly correlated between the left and right somatomotor ROIs ($P < 0.05$). Critically, none of the oculomotor ROIs was significantly correlated with either the left or right somatomotor cortex (all $P > 0.1$). An additional analysis tested the degree to which correlation maps derived from various ROIs were spatially similar⁸ (Supplementary Fig. 3b). All correlation maps corresponding to ROIs within the oculomotor system were spatially correlated (two-tailed t -test, all $P < 0.05$) as were the maps corresponding to the left and right SMC ($P < 0.05$). Notably, none of the correlation maps derived from the oculomotor ROIs was spatially correlated with the maps derived from the somatomotor ROIs (all $P > 0.1$).

To examine the similarity between task-evoked and spontaneous BOLD correlation patterns, we compared the map of spontaneous oculomotor system correlations to a map of BOLD responses evoked during performance of a saccadic eye movement task in two awake

monkeys (data from ref. 9). Figure 2a, b demonstrates that the pattern of saccade task-evoked activations resembles the distribution of spontaneous BOLD correlations in the oculomotor system, which is confirmed by significant spatial correlations between the saccade task-evoked response map and the spontaneous BOLD correlation maps derived from each oculomotor ROI (two-tailed t -test, all $P < 10^{-4}$). These results suggest that the systems showing coherent spontaneous BOLD fluctuations under anaesthesia are similar to systems commonly engaged concurrently during task performance in awake animals.

To evaluate whether the correlation structure of spontaneous BOLD fluctuations relates to the underlying anatomical circuitry, we compared the present oculomotor system correlation maps to a previous study¹⁰ of retrograde tracer injections into area LIP. Figure 2c shows the density of retrogradely labelled cortical cells (averaged across three monkeys) projected onto the cortical surface. A marked similarity is evident in panels a and c of Fig. 2, which is confirmed by significant spatial correlations between the retrograde tracer map and

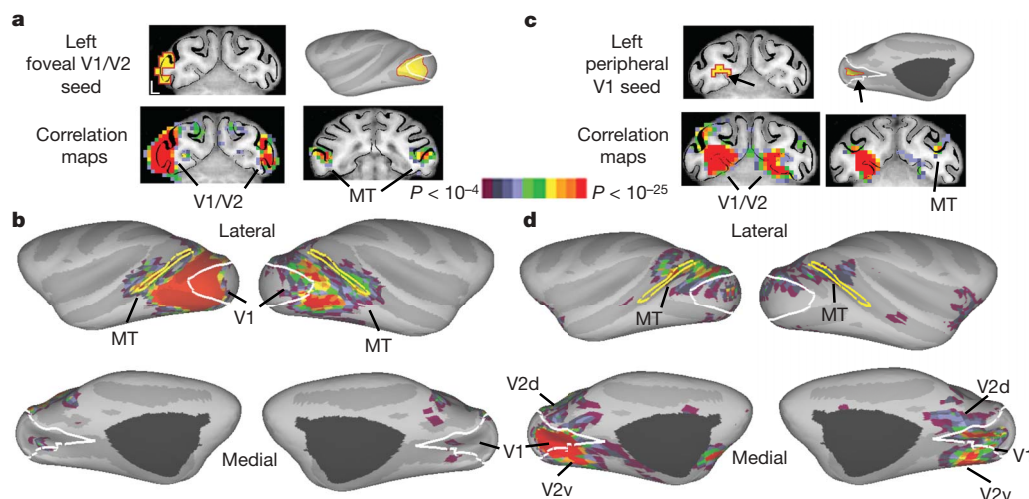


Figure 3 | Spontaneous BOLD correlations are topographically organized in the visual cortex of anaesthetized monkeys ($N = 8$). **a**, Left foveal V1/V2 ROI (yellow) on a coronal slice (upper left) and inflated surface (upper right). Correlation maps on coronal slices demonstrate correlations in contralateral foveal V1 and V2 (lower left) and bilaterally in and near area MT (lower right). **b**, Correlations displayed on inflated views of left and right hemispheres are stronger in regions representing the fovea compared with the periphery in visual areas V1, V2 and MT, based on published

topographic maps^{11,29,30}. **c**, Left peripheral V1 ROI (yellow, upper panels). Coronal slice views demonstrate correlations in contralateral peripheral V1 and V2 (lower left) and bilaterally in area MT (lower right). **d**, Correlations displayed on inflated views are stronger in peripheral compared with foveal representations in V1, V2 and MT. See Supplementary Fig. 7 for analogous results for the right hemisphere seed ROIs. Data are available at <http://sumsdb.wustl.edu/sums/directory.do?id=6602646>.

the spontaneous BOLD correlation maps derived from each oculo-motor ROI (two-tailed t -test, all $P < 10^{-4}$).

To test further the relation to known anatomical circuitry, we analysed correlation patterns associated with retinotopically restricted regions of visual cortex. A left foveal V1/V2 ROI revealed a marked pattern of bilateral correlations in the superior temporal sulcus in and near ventral MT, where the fovea is represented¹¹ (Fig. 3a, b). Contralateral correlations were stronger in V2 than in V1 and in the foveal representation compared with the visual periphery. In contrast, a peripheral V1 ROI (within the calcarine sulcus) revealed bilateral correlations that were stronger in and near dorsal MT (representing the periphery) than in ventral MT and were stronger in peripheral compared with foveal V1/V2 (Fig. 3c, d). Thus, the correlation structure of spontaneous BOLD fluctuations in visual cortex exhibits sub-areal, retinotopic organization. Given the absence of direct inter-hemispheric connections involving V1 away from the vertical meridian¹², the correlations in contralateral V1 must be sustained by polysynaptic pathways that respect visual retinotopy along the eccentricity axis. Additional discussion of relations between anatomical and functional connectivity is given in Supplementary Note 1.

The above results concern systems that are well known in both monkeys and humans. However, there exists a set of regions that is commonly deactivated during attention-demanding cognitive tasks (the so-called 'default' system¹³) that, heretofore, has been described only in humans. The default system is posited to support higher mental faculties including understanding others' mental states¹⁴, self-referential behaviour¹⁵, moral reasoning¹⁶, recollection¹⁷ and imagining the future¹⁸. Many of these behaviours have been proposed to be uniquely human¹⁹. Although anatomical tract tracing results suggest that elements of a similar system may exist in the monkey²⁰, some components of the human default system reside in areas (specifically, Brodmann areas 39 and 40) that have no clear homologue in the monkey brain^{21,22}.

In humans, spontaneous fluctuations within the default system are robustly correlated with the posterior cingulate/precuneus cortex (pC/PCC)^{3,4}. We examined the spontaneous BOLD correlations associated with the monkey pC/PCC (defined anatomically) to determine whether a similar system exists in the monkey. Regions significantly correlated with the pC/PCC in anaesthetized monkeys included dorsal medial prefrontal cortex, lateral temporoparietal cortex (including area 7a and superior temporal gyrus) and posterior parahippocampal cortex (Fig. 4a). Figure 4b shows the analogous result obtained in ten humans scanned in a resting state (see Methods). The resemblance between the monkey and human pC/PCC correlation maps suggests that many elements of the default system may be conserved across primate species. More speculatively, the similarity of the pC/PCC correlation with lateral temporoparietal cortex in both species suggests a potential homologue of human areas 39 and 40 in the macaque. Future research may reveal whether the systems illustrated in Fig. 4 have similar functions across species (see Supplementary Note 2).

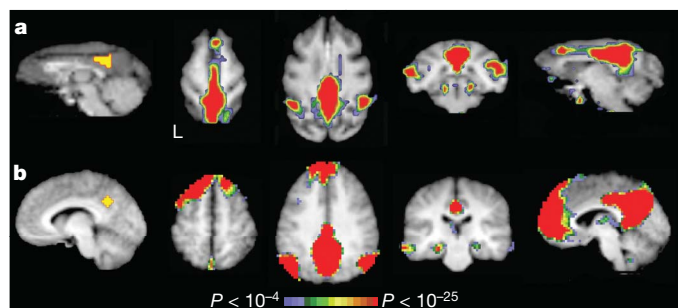


Figure 4 | Significant voxel-wise correlations of the pC/PCC in monkey and human. The seed ROI used in each species is shown on the left. **a**, Data from eight anaesthetized monkeys. **b**, Data from ten resting humans.

Our results demonstrate that cortical systems previously associated with the performance of sensory, motor, and/or cognitive tasks are manifest in the correlation structure of spontaneous BOLD fluctuations observed in the absence of normal perception or behaviour. Thus, coherent spontaneous BOLD fluctuations cannot exclusively be a reflection of conscious mental activity⁷, but may reflect a more fundamental or intrinsic property of functional brain organization. These findings are consistent with the perspective that the brain is governed primarily by internal dynamics^{23–28}. Furthermore, our findings introduce a new approach for comparing the functional architecture of the brain across species.

METHODS SUMMARY

fMRI was acquired in 11 anaesthetized adult macaque monkeys (eight *Macaca fascicularis*; three *Macaca mulatta*). Protocols for anaesthesia and magnetic resonance imaging were reviewed and approved by the Animal Studies Committees of Washington University. Each monkey was initially chemically restrained with ketamine (10 mg kg⁻¹ for *M. mulatta*; 15 mg kg⁻¹ for *M. fascicularis*) and atropine (0.05 mg kg⁻¹) to decrease bronchial and salivary secretions. After intubation, anaesthesia was maintained for the duration of the scan with isoflurane (0.8–1.5%). The anaesthetic level was adjusted to eliminate responses (somatic movement or cardiac rate change) to toe pinches while keeping the heart rate above 70 beats per minute. Corneal reflexes were absent at all times. Neuromuscular blockade was not used. Four monkeys were scanned with simultaneous EEG and fMRI; eight monkeys were scanned without simultaneous EEG. One monkey was used in both studies. Between 6 and 23 BOLD runs of 300 functional volumes (~15 min run⁻¹) were acquired in each monkey. Pre-processing of fMRI data and analysis of spontaneous BOLD correlations was performed as previously described^{3,6,8}. The BOLD time course in functionally and anatomically defined ROIs was correlated with the time course in other ROIs and with the time course at each brain voxel to create spontaneous BOLD fluctuation correlation maps. Surface representations were created using Caret software (<http://brainmap.wustl.edu/caret/>). Correlation maps representing the oculo-motor system were compared to maps of saccadic eye-movement-evoked fMRI responses⁹, anatomical connectivity maps determined by ventral LIP retrograde tracer injections¹⁰, and correlation maps obtained in ten adult humans performing continuous fixation^{3,6,8}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 November 2006; accepted 14 March 2007.

1. Raichle, M. E. & Mintun, M. A. Brain work and brain imaging. *Annu. Rev. Neurosci.* **29**, 449–476 (2006).
2. Biswal, B., Yetkin, F. Z., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**, 537–541 (1995).
3. Fox, M. D. *et al.* The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl Acad. Sci. USA* **102**, 9673–9678 (2005).
4. Greicius, M. D., Krasnow, B., Reiss, A. L. & Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl Acad. Sci. USA* **100**, 253–258 (2003).
5. Damoiseaux, J. S. *et al.* Consistent resting-state networks across healthy subjects. *Proc. Natl Acad. Sci. USA* **103**, 13848–13853 (2006).
6. Vincent, J. L. *et al.* Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J. Neurophysiol.* **96**, 3517–3531 (2006).
7. Morcom, A. M. & Fletcher, P. C. Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage* (in the press).
8. Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L. & Raichle, M. E. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl Acad. Sci. USA* **103**, 10046–10051 (2006).
9. Baker, J. T., Patel, G. H., Corbetta, M. & Snyder, L. H. Distribution of activity across the monkey cerebral cortical surface, thalamus and midbrain during rapid, visually guided saccades. *Cereb. Cortex* **16**, 447–459 (2006).
10. Lewis, J. W. & Van Essen, D. C. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137 (2000).
11. Maunsell, J. H. & Van Essen, D. C. Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. *J. Comp. Neurol.* **266**, 535–555 (1987).
12. Van Essen, D. C., Newsome, W. T. & Bixby, J. L. The pattern of interhemispheric connections and its relationship to extrastriate visual areas in the macaque monkey. *J. Neurosci.* **2**, 265–283 (1982).

13. Raichle, M. E. *et al.* A default mode of brain function. *Proc. Natl Acad. Sci. USA* **98**, 676–682 (2001).
14. Saxe, R. & Kanwisher, N. People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *Neuroimage* **19**, 1835–1842 (2003).
15. Vogeley, K. & Fink, G. R. Neural correlates of the first-person-perspective. *Trends Cogn. Sci.* **7**, 38–42 (2003).
16. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. An fMRI investigation of emotional engagement in moral judgment. *Science* **293**, 2105–2108 (2001).
17. Wagner, A. D., Shannon, B. J., Kahn, I. & Buckner, R. L. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* **9**, 445–453 (2005).
18. Addis, D. R., Wong, A. T. & Schacter, D. L. Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* **45**, 1363–1377 (2007).
19. Tulving, E. in *The Missing Link in Cognition: Origins of Self-Reflective Consciousness* (eds Terrace, H. S. & Metcalfe, J.) 3–56 (Oxford Univ. Press, New York, 2005).
20. Suzuki, W. A. & Amaral, D. G. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J. Comp. Neurol.* **350**, 497–533 (1994).
21. Karnath, H.-O. New insights into the functions of the superior temporal cortex. *Nature Rev. Neurosci.* **2**, 568–576 (2001).
22. Brodmann, K. *Localisation in the Cerebral Cortex* (Smith-Gordon, London, 1994).
23. MacLean, J. N., Watson, B. O., Aaron, G. B. & Yuste, R. Internal dynamics determine the cortical response to thalamic stimulation. *Neuron* **48**, 811–823 (2005).
24. Llinas, R. *I of the Vortex: From Neurons to Self* (MIT Press, Cambridge, Massachusetts, 2001).
25. Buzsaki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).
26. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. Spontaneously emerging cortical representations of visual attributes. *Nature* **425**, 954–956 (2003).
27. Raichle, M. E. Neuroscience. The brain's dark energy. *Science* **314**, 1249–1250 (2006).
28. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **431**, 573–578 (2004).
29. Van Essen, D. C., Newsome, W. T. & Maunsell, J. H. The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Res.* **24**, 429–448 (1984).
30. Gattass, R., Sousa, A. P. & Gross, C. G. Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.* **8**, 1831–1845 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. L. Price, J. S. Perlmutter and G. C. DeAngelis for discussion and for allowing us to scan their monkeys; L. J. Larson-Prior for providing the human data; J. Harwell for Caret software enhancements used in data analysis; and K. J. Black for providing the monkey atlas target. Grants from the US National Institute of Health, US National Science Foundation, Washington University Silvio Conte Center, and Mallinckrodt Institute of Radiology supported these studies.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.E.R. (marc@npg.wustl.edu).

METHODS

Subjects and data acquisition: monkeys. Eleven healthy adult macaque monkeys (eight *Macaca fascicularis*; three *Macaca mulatta*) were used in this study. Protocols for anaesthesia and magnetic resonance imaging were reviewed and approved by the Animal Studies Committee of Washington University. Each animal was initially chemically restrained with ketamine (10 mg kg^{-1} for *M. mulatta*; 15 mg kg^{-1} for *M. fascicularis*) and administered atropine (0.05 mg kg^{-1}) to decrease bronchial and salivary secretions. After intubation, anaesthesia was maintained for the duration of the scan with isoflurane (0.8–1.5%). The anaesthetic level was adjusted to eliminate responses (somatic movement or cardiac rate change) to toe pinches while keeping the heart rate above 70 beats per minute. Corneal reflexes were consistently absent at all times. Neuromuscular blockade was not used. Four monkeys were scanned with simultaneous EEG and fMRI; eight monkeys were scanned without simultaneous EEG. One monkey participated in both studies.

The four monkeys in the EEG–fMRI experiments were initially maintained at 0.8–1.1% isoflurane for between 8 and 15 BOLD runs and then at the higher level of 1.2–1.5% for between 8 and 22 runs. Electrophysiological depth of anaesthesia was monitored throughout scanning by an expert electroencephalographer (J.M.Z.) and graded according to the EEG pattern: continuous low frequency activity was defined as light anaesthesia whereas a clear burst suppression pattern was defined as deep anaesthesia.

Magnetic resonance scanning was performed at the Washington University School of Medicine using a 3T Allegra scanner (Siemens). fMRI data were acquired using a gradient-echo echo-planar sequence sensitive to BOLD contrast (volume repetition time (TR) = 3.02 s, T2* evolution time (TE) = 25 ms, flip angle = 90° , 1.5 mm^2 in plane resolution, slice thickness = 1.6 mm). Two monkeys were scanned using a surface coil; the remaining monkeys were scanned using a volumetric coil (both coils from Primatrix). In three monkeys, each whole-brain volume consisted of 52 coronal slices. In the remaining monkeys, each volume consisted of 52 sagittal slices. In all monkeys, the slices were acquired using contiguous, interleaved acquisition. Between 6 and 23 BOLD runs of 300 functional volumes (300 volumes are approximately 15 min) were acquired in each monkey. Sagittal, high-resolution, T1-weighted, magnetization-prepared rapid gradient echo (MP-RAGE) structural images were also acquired (TR = 1.85 s; TE = 3.93 ms, flip angle = 7° , 0.5 mm^3 voxels) and used to align the functional data to a monkey atlas^{31,32}.

Electroencephalograms were recorded using a MagLink (Compumedics Neuroscan) system equipped with a 24-bit Synamps/2 DC amplifier. The mastoid, VEOG and EKG electrodes (sintered Ag/AgCl) of the MagLink cap were attached to the head of the monkey using collodion. The right and left posterior electrodes were placed 2 cm anterior and lateral to theinion with the right and left anterior electrodes 3 cm anterior to the posterior electrodes. The reference electrode was placed on the midline 3 cm anterior to the anterior electrodes and the ground was placed 2 cm to the left of the reference electrode. All electrode leads were carbon fibre in series with a current limiting resistor for subject safety. Electrode impedances were kept below 10 k Ω . Scan 4.3 Acquire (Compumedics Neuroscan) was used to record EEG both in referential and bipolar modes at a 20 KHz sampling rate (rate needed for subsequent gradient artefact reduction). In general, EEG recorded in a bipolar configuration was of higher quality. EKG was recorded using a bipolar derivation (Millenium monitor, InVivo Research). The electrophysiological data were band-pass filtered (1–20 Hz, 48 dB roll-off) and magnetic resonance gradient artefact was effectively eliminated using Scan 4.3 Edit (Compumedics Neuroscan). The gradient-artefact-reduced electrophysiological data were then decimated to 500 Hz. Ballistocardiogram was reduced using in-house software³³.

Subjects and data acquisition: humans. Ten normal right-handed human participants (mean age = 23.2 ± 2.6 yr; four males) underwent three 5.5-min fMRI scans (Siemens Allegra) during maintained visual fixation. Additional technical details are given in previous reports in which these data were used to study spontaneous BOLD fluctuations^{3,6,8}.

Processing of imaging data. For both species, the fMRI pre-processing steps included: first, compensation of systematic, slice-dependent time shifts; second, elimination of systematic odd-even slice intensity differences due to interleaved acquisition; and third, rigid-body correction for inter-volume head motion within and across runs. Step three provided a record of head position within and across all fMRI runs. Each fMRI run was intensity scaled (one multiplicative constant over all voxels and functional volumes) to yield a whole brain mode value of 1,000 (not counting the first four functional volumes)³⁴. Atlas registration was achieved by computing affine transforms connecting the fMRI run first functional volume (averaged over all runs after cross-run realignment) with the average T1-weighted structural images³⁴. Our macaque atlas representative template includes MP-RAGE data from 11 normal cynomolgus monkeys and was made to conform to the 2000 Martin and Bowden atlas³⁵ (see <http://www.purl.org/net/kbmd/cyno>). Our human atlas representative template includes

MP-RAGE data from 12 normal individuals and was made to conform to the 1988 Talairach atlas³⁶ according to the method of ref. 37.

To prepare the BOLD data for the present main analyses, each fMRI run was transformed to atlas space. The human BOLD data were resampled to 3-mm cubic voxels. The monkey BOLD data were resampled to 1.5-mm cubic voxels. This step combined movement correction within and across runs and atlas transformation in a single resampling.

Several processing steps were used to condition the functional data for analysis of voxel-based correlations as described previously^{3,6}. Monkey and human data were temporally filtered to retain frequencies in the $0.0025 < f < 0.05$ Hz and $0.009 < f < 0.08$ Hz bands, respectively. The monkey and human data were spatially smoothed with 3 mm and 6 mm full-width at half-maximum (FWHM) gaussian kernels, respectively. Several sources of spurious variance were removed from the time series by regression of nuisance variables and their temporal derivatives: (1) six parameters obtained by rigid-body correction of head motion; (2) the whole-brain signal averaged over a fixed region in atlas space. The humans also had the signal from a ventricular region of interest and a region centred in the white matter removed. This regression procedure removes fluctuations unlikely to represent regionally specific correlations of neuronal origin.

Creation of ROIs. Monkey oculomotor ROIs were defined from a previous study of saccadic eye movements⁹. Activation maps were assigned a threshold at a significance level of $P < 0.05$ (fixed effects analysis). ROIs included all contiguous voxels surrounding the most significantly activated peak around left and right LIP and FEF. The LIP ROIs included activated voxels on the lateral bank of the intraparietal sulcus. The FEF ROIs included activated voxels on the anterior bank of the arcuate sulcus. In addition, left and right somatomotor cortex ROIs were defined around Brodmann area 4 according to the partitioning scheme of ref. 38, registered from F99 atlas to the present F6 cortical surface (see ‘Surface-based mapping’ below). The monkey posterior cingulate/precuneus cortex (pC/PCC) ROI was anatomically defined according to the atlas of ref. 35. Similarly, the ROIs for visual topographic subregions were delineated using the partitioning scheme of ref. 38 for areas V1 and V2 and known topographic organization of V1 and V2²⁹. The oculomotor, somatomotor and visual ROIs were drawn on the atlas cortical surface, and then mapped to voxels within 1.5 mm of the specified surface nodes.

The human posterior cingulate/precuneus was defined as a 12-mm-diameter sphere centred on a peak focus from a meta-analysis of default activity (Talairach coordinates: $-7, -53, 34$)³⁹.

Regional correlations. Pearson correlation coefficient between region pairs were computed as

$$r_{xy} = \frac{(1/T) \sum_{t=1}^T [x(t) - \bar{x}] \cdot [y(t) - \bar{y}]}{s_x s_y}$$

where $x(t)$ and $y(t)$ are the regional time courses with means \bar{x} and \bar{y} , respectively, and standard deviations s_x and s_y , respectively. The summation limit, T , corresponds to the total number of time points in multiple, concatenated BOLD runs, excluding the first four (pre-magnetization steady state) functional volumes of each run.

Correlation mapping. For each subject and each ROI, correlation maps were computed as previously described^{2,3,6} by correlating a selected regional time course, $x(t)$, against all other voxels in the brain. Thus,

$$r_x(v) = \frac{(1/T) \sum_{t=1}^T [x(t) - \bar{x}] \cdot [I(v, t) - \bar{I}(v)]}{s_x s_{I(v)}}$$

where $I(v, t)$ represents the time course at locus v and time t with mean $\bar{I}(v, t)$ and standard deviation $s_{I(v)}$. The coefficient $r_x(v)$ is the correlation between the time course at voxel v and the regional time course, $x(t)$.

Application of Fisher's z transform⁴⁰

$$z(v) = 0.5 \ln \left(\frac{1 + r(v)}{1 - r(v)} \right)$$

yielded maps with values at each voxel that theoretically are nearly normally distributed over the population of subjects.

Fixed effects significance maps were created for each individual by converting $z(v)$ maps to Z score maps (that is, zero mean, unit variance, gaussian distributions under the null hypothesis of no correlation). Thus, $z(v)$ was divided by the square root of the theoretical variance, computed as $1/(n-3)$ where n is the degrees of freedom. To account for autocorrelation in the BOLD signal according to Bartlett's theory, n was taken as the total number of time points (functional volumes) used to compute $z(v)$, divided by the time integral of the square of the lagged autocorrelation function⁴¹. Z score maps were combined across subjects

using a fixed-effects analysis (sum and divide by the square root of number of subjects) to generate group Z score maps. The group Z score maps were converted to maps of probability for display in Figs 3 and 4 and Supplementary Figs 1, 2 and 7. All statistical maps were corrected for multiple comparisons using a stringent Bonferroni correction (proportional to total number of brain voxels) and a threshold set at $P < 10^{-4}$. These group statistical maps were used to compute conjunction maps.

Conjunction maps. Conjunction maps were used to visualize voxels consistently correlated with all ROIs within a spontaneously emerging functional system (for example, the four oculomotor ROIs). Conjunction maps were computed on the basis of the Bonferroni corrected group statistical maps. Thus, at each voxel, the number of group statistical maps that were significant at the $P < 10^{-4}$ level were counted. Voxels meeting this criterion for 3 out of 4 oculomotor ROIs are displayed in Fig. 2a and Supplementary Fig. 1e. A somatomotor system conjunction map is shown in Supplementary Fig. 2c.

Conjunction maps were also used in Supplementary Fig. 5 to illustrate voxels consistently correlated with multiple ROIs within the oculomotor and somatomotor systems during anaesthesia. Descriptive rather than hypothesis testing statistics were used to facilitate comparisons across levels of anaesthesia. Thus, these maps were constructed by voxel-wise counting of the number of group $r(v)$ maps with mean correlation (r) ≥ 0.07 .

Evaluation of map similarity: spatial correlation. Similarity of spatial distribution was evaluated for several types of maps (spontaneous BOLD correlation maps, task-evoked BOLD response maps, and anatomical connectivity maps) by computing the standard Pearson correlation coefficient over all voxels within the brain⁸. To evaluate similarity of BOLD fluctuation correlation maps, the Pearson formula was applied after Fisher's transformation (that is, on the basis of $z(v)$). Thus, for correlation maps corresponding to ROIs x and y ,

$$R_{xy} = \frac{\sum_{v=1}^V [z_x(v) - \bar{z}_x] \cdot [z_y(v) - \bar{z}_y]}{\sqrt{\sum_{v=1}^V [z_x(v) - \bar{z}_x]^2} \cdot \sqrt{\sum_{v=1}^V [z_y(v) - \bar{z}_y]^2}}$$

where all summations are taken over all voxels within the brain. The capitalized ' R ' denotes the spatial correlation coefficient whereas the lowercase ' r ' is used for temporal correlation. To generate the results shown in Supplementary Fig. 3b, the spatial correlation coefficient for each pair of $z(v)$ maps was first computed for all individuals. The individual R_{xy} values then were converted to Fisher z values and averaged. Statistical significance was tested using a two-tailed t -test over subjects (random effects analysis) against the null hypothesis of no spatial correlation.

Comparison of spontaneous, evoked and anatomical connectivity patterns. The distribution of BOLD responses evoked by performance of a saccadic eye movement task was obtained in the form of a Z score map from a previous fMRI study of awake behaving monkeys⁹. The similarity between this map and maps of coherent spontaneous activity were independently evaluated for several ROIs (left and right FEF, LIP and somatomotor cortex) using the above-described spatial correlation coefficient. The strategy used to combine map similarity measures across subjects and to evaluate statistical significance was as described above in connection with evaluating similarity of correlation maps. Thus, the spatial correlation coefficient was computed in individuals and converted to Fisher's z . Statistical significance was determined at the group level using a two-tailed t -test against the null hypothesis of no spatial correlation.

Ventral LIP anatomical connectivity data¹⁰ (cases B, C and D) were downloaded from the Surface Management Systems DataBase (<http://sumsdb.wustl.edu/sums/directory.do?id=679531>). These monkeys had retrograde tracer injections (Fast Blue, Sigma, 5% aq.) into ventral LIP. At least 50% of tracer uptake was within ventral LIP; some tracer was also present in the lateral ventral intraparietal area (VIP) in two of the monkeys. For each monkey, the Caret surface metric file containing the density of retrogradely labelled cells was exported as a volume with the same resolution as the atlas-transformed fMRI data ($1.5 \times 1.5 \times 1.5$ mm).

Areas consistently labelled with tracer in all three monkeys were identified by counting, at each voxel, the number of cases containing label. Voxels meeting this criterion in three out of three cases were used as mask for an image of the voxel-wise density of cells labelled by the tracer (averaged over the three cases). This image was re-projected onto a cortical surface representation for display (Fig. 2c). Similarity of LIP anatomical connectivity to spontaneous BOLD correlation maps was independently evaluated for several correlation maps (corresponding to left and right FEF, LIP and somatomotor cortex ROIs) exactly as described above in connection with the saccade task-evoked response map.

Surface-based mapping. To create a surface map of the monkey cortex, T1-weighted MP-RAGE anatomical images of six monkeys (*M. fascicularis*) were

mutually co-registered and averaged using a previously described iterative scheme^{31,42} and a reference target³² representing the atlas of ref. 35 (see <http://www.purl.org/net/kbmd/cyno>). This image then was segmented for the creation of a flattened cortical surface map using the Caret software suite⁴³ to create the 'F6' macaque cortical surface atlas. A nonlinear landmark-driven procedure was used to register the cortical surface to a surface-based monkey brain atlas (F99UA1, <http://sumsdb.wustl.edu/sums/directory.do?id=679531>). Cortical area boundaries established from a previous study were used to assign correlations to specific cortical regions for descriptive purposes³⁸.

The functional correlation maps in Fig. 3 were projected from volume data to F6 atlas surface using the Caret 'average voxel' method with a 1.5-mm averaging radius. This extent was needed in order to capture significant functional data that were slightly misaligned with the structural MRI volume and the atlas surface owing to imperfect registration of the EPI data to the underlying structural MRI. Consequently, the atlas cortical surface maps are blurred as a result of this procedure as well as the initial pre-processing (spatial smoothing of 3 mm FWHM).

- Black, K. J., Koller, J. M., Snyder, A. Z. & Perlmuter, J. S. Atlas template images for nonhuman primate neuroimaging: baboon and macaque. *Methods Enzymol.* **385**, 91–102 (2004).
- Black, K. J., Koller, J. M. & Perlmuter, J. S. Template images for neuroimaging in *Macaca fascicularis*. Program No. 454.18. Abstract Viewer and Itinerary Planner (<http://www.purl.org/net/kbmd/cyno>) (Society for Neuroscience, Washington DC, 2005).
- Vincent, J. L., Larson-Prior, L. J., Zempel, J. M. & Snyder, A. Z. Moving GLM ballistocardiogram artifact reduction for EEG data acquired simultaneously with fMRI. *Clin. Neurophysiol.* **118**, 981–998 (2007).
- Ojemann, J. G. et al. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage* **6**, 156–167 (1997).
- Martin, R. F. & Bowden, D. M. *Primate Brain Maps: Structure of the Macaque Brain* (Elsevier, Amsterdam, 2000).
- Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme Medical Publishers, New York, 1988).
- Lancaster, J. L. et al. A modality-independent approach to spatial normalization of tomographic images of the human brain. *Hum. Brain Mapp.* **3**, 209–223 (1995).
- Lewis, J. W. & Van Essen, D. C. Mapping of architectonic subdivisions in the macaque monkey, with emphasis on parieto-occipital cortex. *J. Comp. Neurol.* **428**, 79–111 (2000).
- Shulman, G. L. et al. Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J. Cogn. Neurosci.* **9**, 648–663 (1997).
- Zar, J. H. *Biostatistical Analysis* (Prentice-Hall, Upper Saddle River, New Jersey, 1996).
- Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications* (Emerson-Adams Press, Boca Raton, 1968).
- Buckner, R. L. et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* **23**, 724–738 (2004).
- Van Essen, D. C. et al. An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assoc.* **8**, 443–459 (2001).

PTC124 targets genetic disorders caused by nonsense mutations

Ellen M. Welch^{1*}, Elisabeth R. Barton^{2*}, Jin Zhuo¹, Yuki Tomizawa¹, Westley J. Friesen¹, Panayiota Trifillis¹, Sergey Paushkin¹, Meenal Patel¹, Christopher R. Trotta¹, Seongwoo Hwang¹, Richard G. Wilde¹, Gary Karp¹, James Takasugi¹, Guangming Chen¹, Stephen Jones¹, Hongyu Ren¹, Young-Choon Moon¹, Donald Corson¹, Anthony A. Turpoff¹, Jeffrey A. Campbell¹, M. Morgan Conn¹, Atiyya Khan¹, Neil G. Almstead¹, Jean Hedrick¹, Anna Mollin¹, Nicole Risher¹, Marla Weetall¹, Shirley Yeh¹, Arthur A. Branstrom¹, Joseph M. Colacino¹, John Babiak¹, William D. Ju¹, Samit Hirawat¹, Valerie J. Northcutt¹, Langdon L. Miller¹, Phyllis Spatrick³, Feng He³, Masataka Kawana², Huisheng Feng², Allan Jacobson³, Stuart W. Peltz¹ & H. Lee Sweeney²

Nonsense mutations promote premature translational termination and cause anywhere from 5–70% of the individual cases of most inherited diseases¹. Studies on nonsense-mediated cystic fibrosis have indicated that boosting specific protein synthesis from <1% to as little as 5% of normal levels may greatly reduce the severity or eliminate the principal manifestations of disease^{2,3}. To address the need for a drug capable of suppressing premature termination, we identified PTC124—a new chemical entity that selectively induces ribosomal readthrough of premature but not normal termination codons. PTC124 activity, optimized using nonsense-containing reporters, promoted dystrophin production in primary muscle cells from humans and *mdx* mice expressing dystrophin nonsense alleles, and rescued striated muscle function in *mdx* mice within 2–8 weeks of drug exposure. PTC124 was well tolerated in animals at plasma exposures substantially in excess of those required for nonsense suppression. The selectivity of PTC124 for premature termination codons, its well characterized activity profile, oral bioavailability and pharmacological properties indicate that this drug may have broad clinical potential for the treatment of a large group of genetic disorders with limited or no therapeutic options.

Nonsense mutations give rise to in-frame UAA, UAG or UGA codons in the messenger RNA coding region, lead to premature translational termination and truncated polypeptide products, and promote mRNA destabilization by nonsense-mediated mRNA decay (NMD)¹. The NMD pathway depends on a set of three conserved factors that modulate both transcript stability and translation termination efficiency^{4,5}. Inactivation of any of these stabilizes nonsense-containing transcripts and promotes nonsense codon readthrough^{6–10}. Such observations indicated that a nonsense-containing mRNA might produce significant amounts of functional protein if either its decay rate or extent of premature termination is altered.

High concentrations of aminoglycosides such as gentamicin promote readthrough of premature nonsense codons in mammalian cells¹¹ and in animal models of nonsense mutation diseases. Gentamicin treatment of the *mdx* mouse—a model of Duchenne muscular dystrophy (DMD)—and of a mouse model of cystic fibrosis led to its evaluation in patients^{12–16}. Treatment of patients harbouring nonsense mutations in the cystic fibrosis transmembrane conductance

regulator (*CFTR*) or dystrophin genes promoted production of the respective missing proteins; however, the lack of potency, the potential renal and otic toxicities, and the need for intravenous or intramuscular gentamicin administration have limited the clinical usefulness of this approach. These proof-of-concept experiments, and the effects of the NMD pathway on mRNA turnover and premature translation termination efficiency, strongly indicated that an orally bioavailable, non-toxic, small-molecule drug that promotes selective and specific readthrough of disease-causing premature termination codons might alleviate the pathologies of nonsense-mediated diseases and we therefore sought an alternative approach.

Two high-throughput screens (comprising ~800,000 low molecular weight compounds) were performed to identify compounds that promoted UGA nonsense suppression. Chemical scaffolds were identified and optimized through extensive medicinal chemistry efforts. Minimally toxic compounds demonstrating UGA readthrough activity were intensively characterized (see Supplementary Information). These analyses identified PTC124 (3-[5-(2-fluorophenyl)-[1,2,4]oxadiazol-3-yl]-benzoic acid; C₁₅H₉FN₂O₃) as a candidate for further development. PTC124 is a 284.24 Da, achiral, 1,2,4-oxadiazole linked to fluorobenzene and benzoic acid rings (Supplementary Figs 1 and 2). The compound has no structural similarity to aminoglycosides or other clinically developed drugs, and its anhydrous free carboxylic acid form, despite having low aqueous solubility (<1 µg ml⁻¹), is orally bioavailable when prepared in aqueous suspension.

PTC124 promoted dose-dependent readthrough of all three nonsense codons in stable cell lines harbouring *LUC-190* nonsense alleles. Levels of suppression correlated inversely with established termination efficiencies^{11,17–19}, with the highest readthrough at UGA, followed by UAG and then UAA (Fig. 1a). Differences in transcript levels do not account for the observed differences in readthrough (Supplementary Fig. 3a). The minimal concentration of PTC124 showing discernable readthrough was 0.01–0.1 µM (2.8–28 ng ml⁻¹), whereas the concentration promoting maximal activity was approximately 3 µM (852 ng ml⁻¹). Because gentamicin is only active at much higher concentrations (Fig. 1b, and data not shown), PTC124 is a more potent nonsense-suppressing agent in this system.

Termination efficiencies are also influenced by the nature of the nucleotide following the nonsense codon (the +1 position)^{18–20}. Like

¹PTC Therapeutics, 100 Corporate Court, South Plainfield, New Jersey 07080, USA. ²Department of Physiology, University of Pennsylvania School of Medicine, 3700 Hamilton Walk, Philadelphia, Pennsylvania 19104, USA. ³Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, Massachusetts 01655, USA.

*These authors contributed equally to this work.

gentamicin¹¹, PTC124 was most active when a pyrimidine (in particular cytosine, C) was located in the +1 position (data not shown). The UGAG termination context was the only efficiently suppressed exception to this trend. Additional experiments with *LUC-190* constructs established that PTC124 did not suppress multiple proximal nonsense codons and that continuous exposure to PTC124 maximizes and maintains suppression activity (Supplementary Fig. 3c–e).

PTC124 promoted suppression of human and mouse nonsense alleles of the dystrophin gene. Drug-treated and untreated primary muscle cell cultures from DMD patients or *mdx* mice were subjected to immunocytochemistry with an antibody recognizing a carboxy-terminal dystrophin epitope. Readthrough of dystrophin mRNA premature nonsense codons was evident in all samples and at all PTC124 concentrations tested. Dystrophin was present at the myofibre membrane, the location critical for maintenance of muscle structural integrity. Consistent with the stable cell line reporter assay (Fig. 1a), the most efficient readthrough was observed in cultures treated with $5 \mu\text{g ml}^{-1}$ ($17 \mu\text{M}$) of PTC124, and there was no further increase at $10 \mu\text{g ml}^{-1}$ (Fig. 2). The ratio of dystrophin:myosin obtained at $5 \mu\text{g ml}^{-1}$ of PTC124 was 40–60% of normal ($n = 2$; Fig. 2 and data not shown). For the *mdx* samples, the dystrophin:myosin ratio was approximately 35% of normal at $10 \mu\text{g ml}^{-1}$ ($n = 8$). Normal myotubes were unaffected by PTC124, and the DMD and *mdx* myotubes had no detectable expression in the absence of drug (Fig. 2).

The ability of PTC124 to promote nonsense suppression was also assessed using *mdx* mice; this analysis required a dosing regimen that would maintain target plasma concentrations of $5\text{--}10 \mu\text{g ml}^{-1}$. Oral feeding of PTC124 by a liquid diet achieved significantly better blood exposure than when administered by intraperitoneal injection (Supplementary Fig. 4). However, to prolong PTC124 exposure during the day (sleep period), the drug was administered intraperitoneally three times per day in addition to oral ingestion by the liquid diet. Using this combination regimen, plasma levels remained $>10 \mu\text{g ml}^{-1}$ at nearly all measured time points (Supplementary Fig. 4).

The functional effects of PTC124 were monitored in *mdx* mice treated with oral, intraperitoneal or combined dosing for 2–8 weeks (during which time there was no significant difference in weight gain

among the treatment and control groups). As shown in Fig. 3a, 4 weeks of PTC124 treatment by any of the 3 treatment regimens partially rescued the functional strength deficit (decreased force per cross-sectional area) characteristic of the *mdx* mouse. Equivalent results were obtained at 2 and 8 weeks (data not shown).

The major functional deficit in dystrophic muscles of *mdx* mice (and, most likely, in DMD patients) is increased susceptibility to contraction-induced injury, especially in muscle that is simultaneously stretched²⁰. This susceptibility results in repeated cycles of degeneration–regeneration, ongoing inflammation and necrosis, with the eventual destruction of muscle. Thus, the best predictor of long-term therapeutic outcome may be protection from contraction-induced injury. The mean percentage drop in force after five eccentric contractions of extensor digitorum longus (EDL) muscles from PTC124-treated or control animals is shown in Fig. 3b. Either intraperitoneal injections or oral dosing alone of PTC124 for 4 weeks resulted in partial protection against contraction-induced injury in the EDL muscles. The effects at 2 and 8 weeks were similar (data not shown). When oral and intraperitoneal dosing were combined (Fig. 3b), a further improvement in protection against contraction-induced injury was observed, such that the decrement in force was not different to that of wild-type (C57) mice.

Elevated levels of creatine kinase are found in the serum of both *mdx* animals and DMD patients. Over 8 weeks, vehicle-treated *mdx* mice showed little change in serum creatine kinase values (data not shown), but those treated with combined oral and intraperitoneal dosing demonstrated significant reductions in serum creatine kinase values by 2 weeks (Fig. 3c), which were maintained for up to 8 weeks. These data corroborate the findings of reduced eccentric contraction injury in the EDLs and suggest that PTC124 leads to decreased muscle fragility.

The functional recovery was associated with dystrophin production, as measured by western blotting. Full-length dystrophin was detected in both C57- and PTC124-treated *mdx* mice (Fig. 3d). *mdx* animals treated with the combined regimen had dystrophin levels approximately 20–25% that of muscles from C57 mice (Fig. 3d). PTC124-treated *mdx* mouse muscles also exhibited increased levels of γ -sarcoglycan (Fig. 3d), consistent with production of dystrophin and stabilization of the dystrophin-associated membrane complex, which is missing in the absence of dystrophin. To confirm the proper membrane localization of dystrophin and associated proteins, striated muscles were subjected to immunohistological analyses. Partial restoration of dystrophin to the membrane was detected in

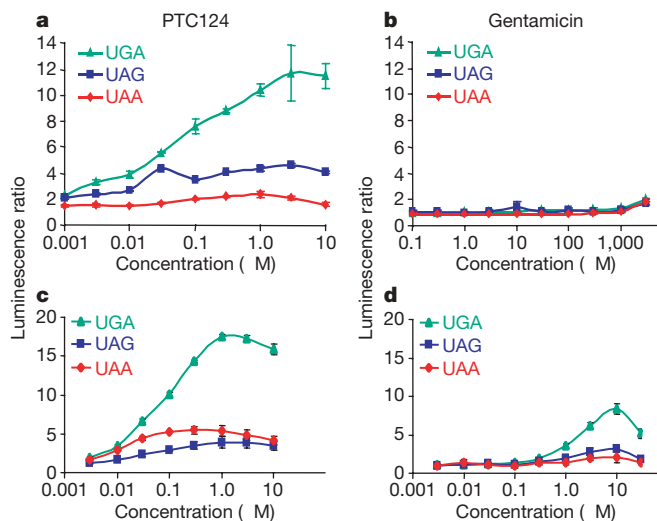


Figure 1 | PTC124 suppresses premature nonsense codons. a, b, Cultured HEK293 cells harbouring UAA, UAG or UGA *LUC-190* nonsense alleles were treated with increasing concentrations of PTC124 (a) or gentamicin (b) for 16 h, and assayed for luciferase activity. c, d, Synthetic *LUC* mRNAs, each harbouring different premature termination codons, were incubated with HeLa cell-free extract supplemented with varying concentrations of PTC124 (c) or gentamicin (d), and assayed for luciferase activity after 4 h. Luminescence ratio, drug-treated:control; error bars (\pm s.d.) are derived from three independent experiments.

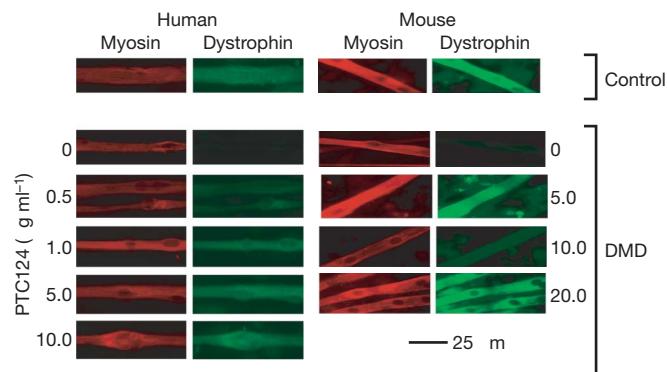


Figure 2 | Full-length dystrophin is produced in PTC124-treated cultured myotubes. Immunohistochemistry of myotubes from primary cell cultures derived from muscle biopsies. Left panel, a non-DMD individual (control) and a patient with a premature UGA codon in exon 28 of the dystrophin gene (DMD). Right panel, wild-type C57 mice (control) and *mdx* mice containing a premature UAA codon in exon 23 of the dystrophin gene (*Dmd*). The effect of adding varying amounts of PTC124 on levels of detectable dystrophin (green) is shown. Myosin (red) levels remain unchanged.

all skeletal muscles examined, including tibialis anterior, diaphragm and cardiac muscle (Fig. 3e). The intensity of fluorescence in drug-treated animals, compared with revertant fibres (white arrow, Fig. 3e), is indicative of lower levels of protein than in wild-type animals, consistent with the data of Fig. 3d.

In principle, nonsense suppression could result from an increase in stop codon readthrough, a decrease in mRNA turnover, or both^{7–10}. To elucidate PTC124's mechanism of action, we monitored its effects on the translation and stability of nonsense-containing mRNAs *in vitro*. PTC124 promoted readthrough at each of the nonsense codons (Fig. 1c), showing maximal activity with UGA, while having no effect on mRNA levels (data not shown). Unlike the stable cell line assays

(Fig. 1a, b), PTC124 did not discriminate significantly between the UAG and UAA mRNAs. PTC124 was a more potent nonsense-suppressing agent than gentamicin (Fig. 1d), and exhibited 4- to 15-fold stimulation of *in vitro* readthrough relative to the controls (Fig. 1c) at levels similar to those in the stable cell reporter assays. These results indicate that PTC124 modulates termination efficiency at premature nonsense codons.

LUC-190 transcripts were also monitored by quantitative PCR with reverse transcription (RT-PCR) in cells treated with PTC124 or cycloheximide, a well characterized translation inhibitor that promotes mRNA stabilization²¹. These experiments showed that nonsense-containing *LUC* mRNA levels increased 11-fold in cycloheximide-treated cells, but were unaffected in PTC124-treated cells (data not shown). These observations were extended to a larger pool of mRNAs—including known NMD substrates²²—by using oligonucleotide microarrays to analyse mRNA expression profiles in HEK293 cells treated for 48 h with or without PTC124 or gentamicin. The averages of six independent comparisons of >54,000-probe sets suggested that very few transcripts deviated from equivalent expression in PTC124-treated versus control cells (Fig. 4a); only 12 transcripts with increased levels and ten with decreased levels were identified in PTC124-treated cells (Supplementary Table 1). None of these are known NMD substrates²² and all 22 were among the lowest expressing cellular mRNAs; that is, they were normally subject to the largest experimental error. No observed differences were detectable by subsequent northern analysis (data not shown). In contrast, six independent comparisons of cells treated or not treated with gentamicin indicated increases in the abundance of 20 transcripts and decreases in the abundance of 11 additional transcripts, including several for which the fold-change was substantial (Fig. 4b; Supplementary Table 2; Supplementary Fig. 5). There was no overlap of transcripts with altered levels in PTC124-treated cells versus gentamicin-treated cells (Supplementary Tables 1 and 2). These results indicate that: (a) the synthesis and stability of few, if any, cellular mRNAs are altered in response to levels of PTC124 that promote nonsense suppression; (b) PTC124 demonstrates little off-target activity at the level of transcription or mRNA stability; and (c) PTC124 and gentamicin have distinct effects on gene expression. Consistent with the latter conclusion, PTC124 does not manifest antibacterial activity (Supplementary Table 3).

Nonsense-suppressing drugs could theoretically promote non-specific readthrough of normal termination codons, but available evidence suggests that normal and premature termination differ mechanistically^{9,10,23}, thereby implying a basis for selectivity. To test directly for the selectivity of PTC124-promoted readthrough, we monitored the accumulation of truncated, full-length and readthrough luciferase polypeptides in drug-treated HEK293 cells, and also assayed for the presence of specific readthrough polypeptides in humans, rats and dogs. Cells harbouring the *LUC-190-CD40* construct (Fig. 4c, d) were treated with PTC124, and luciferase immunoprecipitated from cell lysates was analysed by western blotting. In the absence of drug, translation of the *LUC-190-CD40* mRNA yielded only a 25 kD amino-terminal truncated luciferase fragment, but treatment of cells with PTC124 led to a dose-dependent accumulation of full-length luciferase (Fig. 4c). Importantly, no luciferase was detected that corresponded in size to the product expected from readthrough of the normal terminator (Fig. 4c). The absence of a detectable readthrough product was unlikely to be a consequence of its instability because the control readthrough protein, generated by replacing the termination codon with CGA, was efficiently expressed in the same cells (Fig. 4c, d). Similar analyses also failed to detect polypeptides corresponding to putative readthrough products in multiple tissues isolated from PTC124-treated human subjects, rats and dogs (Supplementary Fig. 6). For example, western blotting analyses of pooled peripheral blood mononuclear cells from subjects treated with 200 mg kg⁻¹ of PTC124 elicited only wild-type $\beta 2$ microglobulin and no additional, longer polypeptides that would

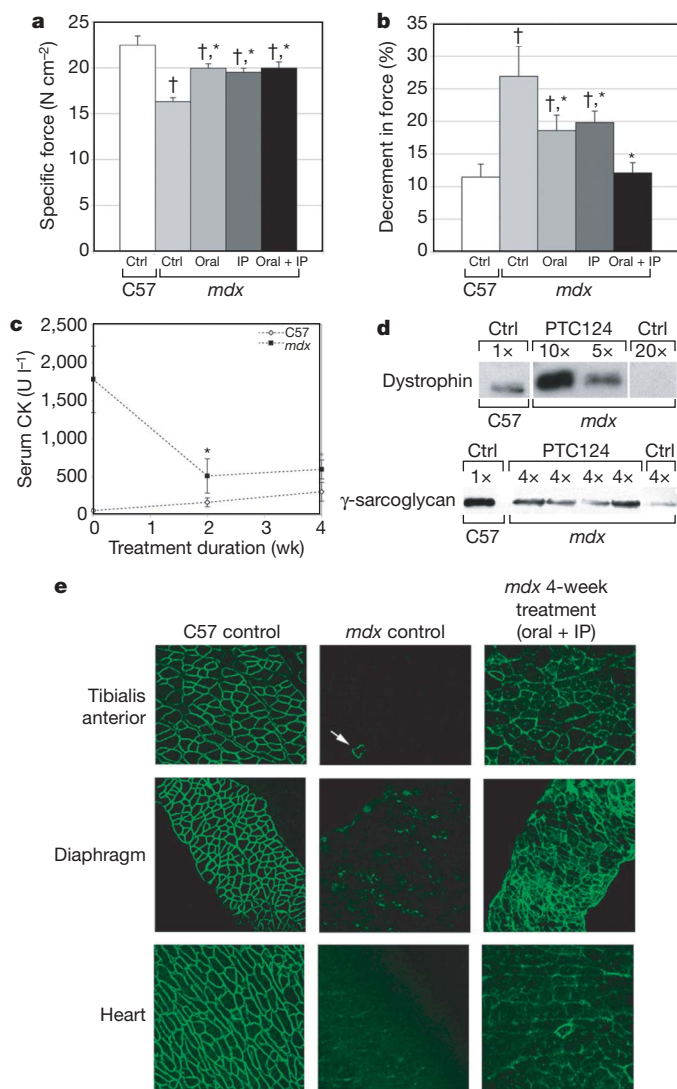
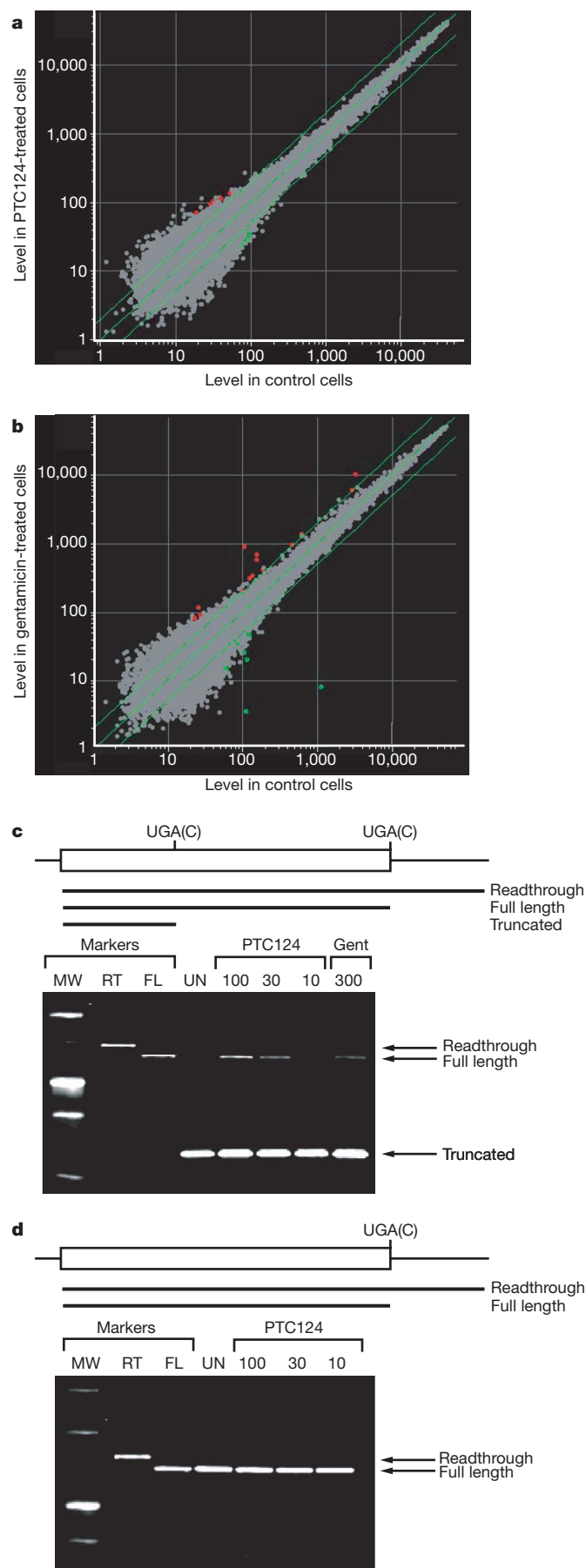


Figure 3 | Rescue of the dystrophic phenotype in muscles of the *mdx* mouse. **a**, The effects of oral dosing, intraperitoneal (IP) injections or combined dosing of PTC124 on the force per cross-sectional area (specific force) of the EDL muscle after 4 weeks of drug treatment. **b**, The effects of 4 weeks of oral dosing, IP injections or combined dosing of PTC124 on preventing loss-of-force production in the EDL muscle following five successive eccentric contractions. **c**, Serum creatine kinase (CK) changes in *mdx* mice after PTC124 dosing. $\dagger P < 0.05$ for comparisons to untreated C57 mice; $* P < 0.05$ for comparisons to untreated *mdx* mice. **d**, Western blot analysis of quadriceps and tibialis anterior muscles for dystrophin and γ -sarcoglycan. Protein load of each lane is indicated relative to the load of C57 wild-type (1 \times). Ctrl, control. **e**, Immunohistochemistry of indicated muscle cross-sections to visualize dystrophin in the tibialis anterior, diaphragm and heart. The white arrow in the tibialis anterior negative control designates a revertant fibre. **a–c**, Data expressed as means \pm s.e.m.; $n = 8$.



have originated from termination codon readthrough (Supplementary Fig. 6c). These results indicate that PTC124 promotes readthrough of premature termination without affecting normal termination, even at drug exposure levels substantially greater than the values achieving maximal activity. Consistent with this conclusion, two-dimensional gel analyses of HEK293 cells incubated with or without PTC124 showed no significant differences in the relative amounts or shapes of the respective spots representing several hundred polypeptides (Supplementary Fig. 7).

Systemic delivery of PTC124 was achieved without any obvious toxicity, consistent with the findings of preclinical safety pharmacology and toxicology studies in rats and dogs²⁴. Given the significance of NMD as a genomic surveillance mechanism²⁵, PTC124's lack of effect on this pathway may be an important component of its safety. The safety of PTC124 may also be related to the observation that its readthrough activity is specific for premature stop codons. Such selectivity might be expected from the apparent mechanistic differences between premature and normal termination²³, but may also be enhanced by the inability of PTC124 to efficiently promote readthrough of the multiple stop codons normally present in mRNA 3'-UTRs²⁶, and by the specific mRNA decay mechanism known to be activated when translation extends into the 3'-UTR^{27,28}.

Clinical trials of PTC124 have been initiated and their successful completion may ultimately allow therapy of subsets of patients in a large and diverse group of genetic disorders for which the primary disease defect is the presence of a nonsense mutation. As such, this approach is among the first to test the model of personalized medicine, in which the focus shifts from treatment of a disease to treatment of a specific genetic defect.

METHODS SUMMARY

PTC124 was identified by a combination of high-throughput screening of a small-molecule chemical library and subsequent lead optimization of compounds exhibiting significant nonsense suppression activity and low toxicity. Optimization protocols selected for high oral bioavailability, lack of *in vitro* off-target activity, *in vivo* safety and suitability for pharmaceutical formulation. Nonsense suppression analyses in the screening and early characterization steps monitored the production of luciferase from *LUC* nonsense-containing mRNAs expressed in HEK293 cells or incubated as synthetic transcripts in cell-free lysates. Suppression of dystrophin nonsense alleles, in patients and in *mdx* mice, was assayed by established procedures, including those monitoring dystrophin accumulation *in vivo* and in cultured myotubes, and those analysing dystrophin function as inferred from the mechanics of isolated mouse EDL muscles. Levels of cellular mRNAs were quantified by high-density microarray analysis and northern blotting, and readthrough of normal termination codons was assessed by western blotting using antibodies targeted to either full-length polypeptides or to putative readthrough-dependent C-terminal extension peptides. Detailed protocols for all assays can be found in Methods.

Figure 4 | PTC124 activity is selective for readthrough of premature translation termination codons. **a, b**, Microarray analyses of mRNA levels in HEK293 cells treated with 5 μ M of PTC124 (**a**) or ~300 μ M of gentamicin (**b**). The relative level of each transcript in the drug-treated samples was normalized to that in the untreated samples and the resulting expression ratios from 5/6 independent replicates were averaged and plotted on a logarithmic scale for pair-wise comparisons. The centre line indicates the line of equivalence and the outer lines indicate a twofold difference in expression. **c, d**, PTC124 does not promote readthrough of normal termination codons. Shown are a cartoon depicting *LUC* reporter constructs, and western blot analysis of *LUC* readthrough protein. Cells harboured *LUC* nonsense (UGAC) or wild-type alleles with 6 \times -histidine and Xpress epitope tags, inserted in frame with the *LUC* coding region, as well as a CD40 3'-UTR. Cultures were treated with PTC124 (100 μ M (28.4 μ g ml⁻¹), lane 5; 10 μ M (2.8 μ g ml⁻¹), lane 6; and 1.0 μ M (0.28 μ g ml⁻¹), lane 7) for 72 h, and luciferase protein was purified and analysed by western blotting. Readthrough marker protein was derived by mutating the normal termination codon to CGA. *M_r*, relative molecular mass marker; RT, readthrough protein marker; FL, full-length protein marker; UN, untreated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 September 2006; accepted 16 March 2007.

Published online 22 April 2007.

- Mendell, J. T. & Dietz, H. C. When the message goes awry: disease-producing mutations that influence mRNA content and performance. *Cell* **107**, 411–414 (2001).
- Kerem, E. Pharmacologic therapy for stop mutations: how much CFTR activity is enough? *Curr. Opin. Pulm. Med.* **10**, 547–552 (2004).
- Ramalho, A. S. *et al.* Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. *Am. J. Respir. Cell Mol. Biol.* **27**, 619–627 (2002).
- Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature Rev. Mol. Cell Biol.* **5**, 89–99 (2004).
- Amrani, N., Sachs, M. S. & Jacobson, A. Early nonsense: mRNA decay solves a translational problem. *Nature Rev. Mol. Cell Biol.* **7**, 415–425 (2006).
- Welch, E. M., Wang, W. & Peltz, S. W. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 467–486 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
- Maderazo, A. B., He, F., Mangus, D. A. & Jacobson, A. Upf1p control of nonsense mRNA translation is regulated by Nmd2p and Upf3p. *Mol. Cell. Biol.* **20**, 4591–4603 (2000).
- Weng, Y., Czaplinski, K. & Peltz, S. W. Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol. Cell. Biol.* **16**, 5477–5490 (1996).
- Weng, Y., Czaplinski, K. & Peltz, S. W. Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Mol. Cell. Biol.* **16**, 5491–5506 (1996).
- Wang, W., Czaplinski, K., Rao, Y. & Peltz, S. W. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *EMBO J.* **20**, 880–890 (2001).
- Manuvakhova, M., Keeling, K. & Bedwell, D. M. Aminoglycoside antibiotics mediate context-dependent suppression of termination codons in a mammalian translation system. *RNA* **6**, 1044–1055 (2000).
- Politano, L. *et al.* Gentamicin administration in Duchenne patients with premature stop codon. Preliminary results. *Acta Myol.* **22**, 15–21 (2003).
- Clancy, J. P. *et al.* Evidence that systemic gentamicin suppresses premature stop mutations in patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **163**, 1683–1692 (2001).
- Wilschanski, M. *et al.* Gentamicin-induced correction of CFTR function in patients with cystic fibrosis and CFTR stop mutations. *N. Engl. J. Med.* **349**, 1433–1441 (2003).
- Wagner, K. R. *et al.* Gentamicin treatment of Duchenne and Becker muscular dystrophy due to nonsense mutations. *Ann. Neurol.* **49**, 706–711 (2001).
- Barton-Davis, E. R., Cordier, L., Shotoruma, D. I., Leland, S. E. & Sweeney, H. L. Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of *mdx* mice. *J. Clin. Invest.* **104**, 375–381 (1999).
- Bonetti, B., Fu, L., Moon, J. & Bedwell, D. M. The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **251**, 334–345 (1995).
- McCaughan, K. K., Brown, C. M., Dalphin, M. E., Berry, M. J. & Tate, W. P. Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proc. Natl Acad. Sci. USA* **92**, 5431–5435 (1995).
- Howard, M. T. *et al.* Sequence specificity of aminoglycoside-induced stop codon readthrough: potential implications for treatment of Duchenne muscular dystrophy. *Ann. Neurol.* **48**, 164–169 (2000).
- Petrof, B. J., Shrager, J. B., Stedman, H. H., Kelly, A. M. & Sweeney, H. L. Dystrophin protects the sarcolemma from stresses developed during muscle contraction. *Proc. Natl Acad. Sci. USA* **90**, 3710–3714 (1993).
- Jacobson, A. & Peltz, S. W. Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu. Rev. Biochem.* **65**, 693–739 (1996).
- Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature Genet.* **36**, 1073–1078 (2004).
- Amrani, N. *et al.* A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**, 112–118 (2004).
- Hirawat, S. *et al.* Safety, tolerability, and pharmacokinetics of PTC124, a nonaminoglycoside nonsense mutation suppressor, following single- and multiple-dose administration to healthy male and female adult volunteers. *J. Clin. Pharm.* **47**, 430–444 (2007).
- He, F. *et al.* Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol. Cell* **12**, 1439–1452 (2003).
- Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137–140 (2001).
- Frischmeyer, P. A. *et al.* An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258–2261 (2002).
- van Hoof, H. A., Frischmeyer, P. A., Dietz, H. C. & Parker, R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**, 2262–2264 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by an STTR grant to A.J. from the NIH and grants from the Muscular Dystrophy Association (USA) and Parent Project Muscular Dystrophy (USA) to H.L.S. We thank L. Cao and T. Komatsu for helpful discussions, G. Elfring for statistical support, D. Minn, X. Kang and S. Gothe for database mining and informatics expertise, and N. Garneau, S. I. Huq, and A. Bhattacharya for technical expertise. We thank K. Donnelly, C. Hirawat and F. P. Nigel for their effort, support and enthusiasm for the project. L. Gold, D. Goeddel and the late R. Swanson provided advice, encouragement and support at the onset of this project, which is gratefully acknowledged. We are grateful to the patients and their families and doctors for their participation in the clinical trial that generated the muscle biopsies and for their commitment during the development of PTC124.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to S.W.P. (speltz@ptcbio.com).

METHODS

LUC constructs. All DNA manipulations used standard procedures²⁹ and were confirmed by sequence analysis. Primary sequences of oligonucleotides used for cloning and mutagenesis are listed in Supplementary Table 7. *LUC-190* constructs were initiated by isolating a DNA fragment containing the encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) from pBS-E-CAT, using *Apal* and *NcoI*, and a fragment containing a β -globin intron from pmCMV-G1-Norm (kindly provided by L. Maquat), using *NcoI* and *BamHI*. The two fragments were inserted into *Apal/BamHI*-digested pBS(SK), the resulting plasmid was digested with *BamHI* and *NotI*, and the firefly *LUC* gene was then inserted into it as a *BamHI-NotI* fragment from p2Luci³⁰. The DNA of interest was transferred to pcDNA3 (-) hygro (Clontech) using the *Apal* and *NotI* sites (pPTC29). Site-directed mutagenesis, using the QuikChange Kit (Stratagene) and oligonucleotides 1–6, was used to introduce TAA, TGA and TAG nonsense mutations at position 190 of the *LUC* coding region (pPTC30, pPTC31 and pPTC32, respectively). To monitor nonsense codon context, the +1 nucleotide was changed to C, T or G by site-directed mutagenesis using oligonucleotides 39–56. To monitor the effect of two or three tandem nonsense codons, *LUC-190* was mutagenized using oligonucleotides 15–18. *LUC-190-CD40* constructs were derived by cloning *LUC190* genes in-frame and downstream of the His6 and Xpress epitopes using plasmid pcDNA4-hygro (Clontech). The CD40 3'-UTR (harbouring 5'-*NotI* and 3'-*Apal* restriction sites prepared using oligonucleotides 7–10) was inserted downstream of the *LUC* coding region and site-directed mutagenesis (using oligonucleotides 11–14) was used to make two mutations creating sense codons in the remaining vector sequence before the CD40 3'-UTR. Translation into the CD40 3'-UTR produces an 84 amino acid (9 kDa) extension of the luciferase protein. To change the termination codon to UGAC, oligonucleotides 57 and 58 were used for site-directed mutagenesis. To prepare plasmid T7-LUC-190, used for the preparation of synthetic mRNA, the pCEL plasmid harbouring the luciferase cDNA³¹ was digested with *HpaI* and *XhoI* to remove the EMCV-Luciferase complementary DNA fragment and the latter DNA was inserted into pBluescriptKS+. This plasmid was digested with *XhoI* for transcription from the T7 promoter. Mutations were introduced using the same oligonucleotides and methods used to construct the stable luciferase reporter cell line constructs.

Assays for readthrough of *LUC* premature termination codons. HEK293 cells growing in medium containing fetal bovine serum (FBS) were stably transfected with *LUC* genes containing premature terminators at codon 190 (and +1 contextual variants thereof) and treated with the indicated concentrations of PTC124, gentamicin or vehicle. Luminescence of each culture relative to the control was quantified by standard procedures. Cell-free translation assays used synthetic *LUC* mRNAs harbouring an EMCV IRES³² and codon 190 premature terminators (with a +1A) prepared using the MegaScript *in vitro* transcription kit (Ambion). The reaction mixtures (20 μ l) contained 16.5 mM HEPES (KOH), pH 7.4, 85 mM potassium acetate, 1.48 mM magnesium acetate, 0.56 mM ATP, 0.075 mM GTP, 18.75 mM creatine phosphate (di-Tris), 1.275 mM dithiothreitol (DTT), amino acids, creatine kinase, HeLa cell cytoplasmic extract (60% of the reaction volume), 100 ng RNA and 0.25% DMSO. Increasing concentrations of PTC124 and gentamicin were added to the *in vitro* reactions, and the amount of luminescence produced was determined after approximately 4 h. Stock solutions of PTC124 (6 mM) were prepared in 100% DMSO (Sigma); gentamicin (sulphate salt, Sigma) stock solutions (1 mM) were prepared in water. Luminescence was determined using a Viewlux CCD imager (Perkin-Elmer). Luminescence data were normalized to that produced with solvent alone (DMSO or water), and the fold suppression over background was calculated as $(\text{PTC124}_{\text{light units}}/\text{DMSO}_{\text{light units}})$ or $(\text{gentamicin}_{\text{light units}}/\text{water}_{\text{light units}})$.

Chemistry. Melting points were determined on an Electrothermal Mel-temp 1201D or Thomas Hoover melting point apparatus, and are uncorrected. ¹H and ¹³C NMR spectra were recorded on a Varian MVX 300 MHz instrument. Mass spectral data of final products were collected from a Waters Micromass ZQ mass spectrometer with an electrospray detector, coupled to a Waters 2795 high-performance liquid chromatograph (LC/MS). Detected were positive ions (ES+) and negative ions (ES-); relative ion intensities are given in per cent. Elemental analyses were obtained by Quantitative Technologies. HPLC methods are as follows: method A—Hewlett-Packard, 254 nm detector; method B—Hewlett-Packard, 280 nm detector; method C—Waters 2795, 254 nm detector.

RNA analysis. Quantitative real-time PCR was performed using primers for *LUC* and dystrophin mRNAs, with the levels of 18S rRNA and GAPDH mRNA serving as normalization factors for the amount of starting material. Northern blot analysis was performed using probes specific to *LUC* mRNA, with the level of GAPDH mRNA serving as a control for loading. High density oligonucleotide arrays (Affymetrix U133⁺, 2.0), containing 54,675 probe sets/chip were used to analyse expression profiles of PTC124-treated, gentamicin-treated

and untreated HEK293 cells. cRNA sample preparation, array hybridization and scanning followed protocols recommended by the manufacturer (<http://www.affymetrix.com/index.affx>). The hybridization intensity (signal) of each transcript was determined using the Affymetrix Microarray Suite 5.0 software package. Intensity values were scaled such that the overall fluorescence intensity of each array was equivalent. Scaling factors for each experimental set varied by less than 20%. Transcript levels specifically enhanced or reduced by drug treatment were identified by applying several stringent standards to the raw data, including: (a) the signal values of a transcript in the treated and untreated samples had to have a minimal level of at least 26 units and a relative change of at least twofold; (b) these changes had to be reproducible in five of the six independent replicate experiments; and (c) these changes had to demonstrate statistically significant *P*-values ≤ 0.05 in a group *t*-test. The GeneSpring (version 7; <http://www.silicongenetics.com>) and GENECLUSTER (version 1.0; <http://www-genome.wi.mit.edu>) software packages were used to identify transcripts that were differentially expressed. Primary data can be found at: <http://jacobsonlab.umassmed.edu/cgi-bin/pubcontents.cgi?pubcontents=2006-PTC>.

Normal luciferase termination codon in stable cell lines. Cells were treated with PTC124 (up to 100 μ M) for 72 h, harvested and lysed, and luciferase was purified using Ni-NTA magnetic agarose beads (Qiagen). Purified proteins were fractionated by PAGE (12.5%) and detected by western blotting, using anti-Xpress antibody (Invitrogen).

Normal termination codons in animals. Sprague-Dawley rats were treated for 14 days with vehicle or PTC124 at doses of 200 to 1,800 mg kg⁻¹ day⁻¹ and male beagles were treated with vehicle or PTC124 at 500 to 1,500 mg kg⁻¹ day⁻¹ (Charles River Laboratories). A 14-day recovery period after the last dose was included. Tissues (heart, kidney, liver, lung, small intestine (jejunum and/or ileum), mandibular salivary glands, thymus and peripheral blood mononuclear cells) were collected between 2 and 8 h after the last dose (that is at the time of expected maximal plasma drug concentration) and again at the end of the recovery period. Tissue pieces (5 mm \times 5 mm) were snap frozen in liquid nitrogen and stored at -70 °C. Frozen tissues were suspended (1:1) with tissue protein extraction reagent (T-PER; Pierce) containing protease inhibitor cocktail (Sigma) diluted 1:100, and homogenized on ice using a Fisher PowerGen 25 homogenizer equipped with an Omni Tip probe. A 100 μ l aliquot was added to 400 μ l of sample buffer²⁹, incubated for 8 min at 100 °C, then transferred to wet ice. Peripheral blood mononuclear cells samples were mixed with 200 μ l of sample buffer, heated to 100 °C for 8 min, and transferred to ice. Homogenized samples (50 μ l aliquots) were pooled by dose level and sex, and fractionated by SDS-PAGE. Western analyses used the ECL Plus chemiluminescent detection kit (Amersham). Primary mouse monoclonal antibodies for α - and β -actin (Oncogene Research Products), cofilin (BD Biosciences), GAPDH (Abcam), U1 snRNP A (Acris) and vimentin (BD Pharmingen), and a goat anti-mouse IgG horseradish peroxidase (HRP) conjugate as secondary antibody (Promega) were used.

Positive controls for animal studies. The readthrough controls for rat vimentin (Mammalian Gene Collection accession number, BC061847), cofilin (BC086533) or U1 snRNP A (BC086331) were amplified using a rat fetal liver cDNA (Marathon, BD Biosciences, Clontech), *Pfu* Turbo DNA Polymerase (Stratagene) and oligonucleotide pairs 19 and 20, 27 and 28, and 23 and 24, respectively. PCR products were purified, digested with *NotI* and *BamHI*, and cloned into pcDNA3.1-hygro (Invitrogen). Mutagenesis of the cloned DNAs, using oligonucleotide pairs 21 and 22, 29 and 30, and 25 and 26 converted the TAA stop codon of vimentin to TAC, the TAG stop codon of cofilin to TGG, and the TGA stop codon of snRNP A to TGG. Dog cofilin (XM_533231) and U1 snRNP A (XM_533663) constructs were prepared using dog kidney cDNA (BioChain) and oesophageal cDNA (BioChain). Oligonucleotide pairs for cloning the cofilin cDNA and then mutagenizing its TGA stop codon to TGG were 35 and 36, and 37 and 38, respectively. For dog U1 snRNP A, PCR amplification used oligonucleotides 31 and 32; mutagenesis to convert the TAG stop codon to TGG used oligonucleotides 33 and 34.

Normal termination codons in healthy volunteers. Peripheral blood mononuclear cells and plasma was obtained from untreated subjects (*n* = 3 at Day 1, Time 0) and at several time points (*n* = 3 per timepoint) relative to the last dose in subjects receiving a single 200 mg kg⁻¹ dose of PTC124. Samples were pooled and primary anti-wild-type-protein antibody was used to probe for wild-type protein and anti-readthrough-protein antibody was used to probe for the readthrough protein (Supplementary Fig. 6c). The antibody to the β 2 microglobulin readthrough protein was produced in rabbits using a peptide corresponding to the predicted translation product (AASWRFEDAAFLGDEFQILLACFLILICLYTYTLCTKCRVIIMLTWT).

Dystrophin expression in primary mouse and human muscle cells. Primary muscle cell cultures of either wild-type (C57) or dystrophin-deficient (*mdx*) mice were obtained as previously described³³. Primary human myoblasts were

derived from muscle biopsies. Two biopsies were obtained from the extensor digitorum brevis (EDB) muscles of DMD patients with premature stop codons in dystrophin exons 28 (UGA) and 42 (UAG), respectively. An additional biopsy was obtained from an unaffected adult human male. Biopsies were rinsed in Ham's F10 medium and minced. Explants were then subjected to a low speed centrifugation, and the pellet was resuspended in medium and transferred to culture dishes with F-10 medium and 20% fetal bovine serum (FBS). On migration of satellite cells from the explants and onto the dish, the cells were harvested and frozen in 95% FBS and 5% DMSO.

To form myotubes, an aliquot of cells was thawed and expanded in growth medium containing 20% FBS, F-10 medium and penicillin–streptomycin. To induce myotube differentiation, the cells were switched to medium containing DMEM, insulin, apo-transferrin and penicillin–streptomycin, and then plated on collagen-coated dishes with Matri-gel³⁴.

At the switch to differentiation medium, PTC 124 was added to normal and DMD cultures in the following doses: 0, 0.5, 1, 5 and 10 $\mu\text{g ml}^{-1}$ of differentiation medium (0, 1.7, 3.4, 17 and 34 μM , respectively). Cultures were allowed to form myotubes and harvested 12 days after switching to differentiation medium. After fixation in 2% formaldehyde, the fixed myotubes were incubated with polyclonal anti-dystrophin antibodies (targeted to the C-terminal region; Abcam 15277) and monoclonal anti-sarcomeric myosin primary antibodies (a gift from Howard Holtzer). The secondary antibodies (Molecular Probes) were conjugated with Alexa 488 for dystrophin and Alexa 586 for myosin. Imaging was performed on a Leica confocal microscope and the ratio of the dystrophin fluorescence intensity to the myosin intensity was obtained to allow estimation of dystrophin levels induced in the DMD and normal cultures. Treated and untreated myotubes appeared healthy, as confirmed by robust myosin staining and the expected presence of striations (Fig. 2).

Assessment of phenotypic rescue of the *mdx* mouse. Isolated whole-muscle mechanics were performed on EDL muscles from drug-treated and control animals using a previously described apparatus⁸. After determining optimum length (L_0) with a series of twitch stimulations at supramaximal voltage, maximum isometric force was measured with 120 Hz pulses delivered for 500 ms. Protection against mechanical injury, induced by a series of 5 eccentric tetanic contractions with stretches of 10% L_0 , was evaluated. Damage was determined as the loss in force between the first and last eccentric contraction. At the end of the mechanical measurements, muscles were blotted and weighed, mounted in embedding compound, then rapidly frozen in melting isopentane. Specific force was calculated from maximum tetanic tension and cross-sectional area of the EDL muscles. Muscle cross-sectional area was determined using previously published methods³³.

An estimate of restoration of the dystrophin glycoprotein complex to the sarcolemma was performed by immunoblotting of muscle homogenates. Samples were partially enriched for the dystrophin-associated glycoprotein complex using wheat germ agglutinin. Antibodies to both dystrophin (C-terminal antibody, Dys2 (Novacastra); rod antibody, Sigma8168 (Sigma-Aldrich)) and gamma-sarcoglycan (NCL-gSarc, Novacastra) were used for these measurements.

The presence and membrane localization of dystrophin was observed by immunohistochemistry on 10-mm frozen sections of striated muscles. The primary antibody for dystrophin was a rabbit polyclonal antibody generated against a synthetic peptide corresponding to the C-terminal of human dystrophin (abcam 15277, Abcam), and diluted 1:1000 in 10% bovine serum albumin, and the secondary antibody (Alexa Fluor 488-conjugated anti-rabbit, Molecular Probes, Invitrogen) was diluted 1:750. Sections were covered with aqueous mounting media containing 4,6-diamidino-2-phenylindole (DAPI; Vectashield, Vector Laboratories) to visualize myonuclei. Stained sections were visualized on a Leica TCS SL confocal microscope (Leica Microsystems), and captured with a digital camera using associated image analysis software (Version 2.61).

All data were expressed as means \pm s.e. A one-way ANOVA was used to compare all variables among groups. Serum CK comparisons used repeated measures test. Significance was indicated if $P < 0.05$ for all comparisons.

Two-dimensional electrophoresis. Duplicate samples of HEK293 cells harbouring *LUC-190* (UGA) were incubated in the presence of 5 μM PTC124 (treated) or 1% DMSO (untreated) for 20 h. The cells were collected, washed twice in phosphate buffered saline (PBS), resuspended in sample buffer (Bio-Rad) and shipped on dry ice to Kendrick Laboratories for two-dimensional electrophoretic analysis³⁵. Isoelectric focusing (pH 3.5–10) was carried out in glass tubes for 20,000 V-hours. One μg of a tropomyosin internal standard was added to each sample. Second dimension SDS slab gel electrophoresis was carried out for approximately 6 h at 25 mA per gel. After electrophoresis, gels were transferred to PVDF paper. Computerized analysis of spot mobility used Phoretix software.

30. Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479–486 (1998).
31. Van Der Velden, V., Kaminski, A., Jackson, R. J. & Belsham, G. J. Defective point mutants of the encephalomyocarditis virus internal ribosome entry site can be complemented *in trans*. *Virology* **214**, 82–90 (1995).
32. Jang, S. K. *et al.* A segment of the 5'-nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during *in vitro* translation. *J. Virol.* **62**, 2636–2643 (1988).
33. Neville, C., Rosenthal, N., McGrew, M., Bogdanova, N. & Hauschka, S. in *Methods in Cell Biology* 52. (eds Sweeney H. L. and Emerson, C.) 85–116 (Academic Press, San Diego, 1998).
34. Sweeney, H. L. & Feng, H. Structure–function analysis of cytoskeletal/contractile proteins in avian myotubes. *Methods Cell Biol.* **52**, 275–282 (1997).
35. O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007–4021 (1975).

29. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989).

LETTERS

Chitin induces accumulation in tissue of innate immune cells associated with allergy

Tiffany A. Reese¹, Hong-Erh Liang¹, Andrew M. Tager², Andrew D. Luster², Nico Van Rooijen³, David Voehringer^{1†} & Richard M. Locksley¹

Allergic and parasitic worm immunity is characterized by infiltration of tissues with interleukin (IL)-4- and IL-13-expressing cells, including T-helper-2 cells, eosinophils and basophils¹. Tissue macrophages assume a distinct phenotype, designated alternatively activated macrophages². Relatively little is known about the factors that trigger these host responses. Chitin, a widespread environmental biopolymer of *N*-acetyl- β -D-glucosamine, provides structural rigidity to fungi, crustaceans, helminths and insects³. Here, we show that chitin induces the accumulation in tissue of IL-4-expressing innate immune cells, including eosinophils and basophils, when given to mice. Tissue infiltration was unaffected by the absence of Toll-like-receptor-mediated lipopolysaccharide recognition but did not occur if the injected chitin was pre-treated with the IL-4- and IL-13-inducible mammalian chitinase, AMCase⁴, or if the chitin was injected into mice that overexpressed AMCase. Chitin mediated alternative macrophage activation *in vivo* and the production of leukotriene B₄, which was required for optimal immune cell recruitment. Chitin is a recognition element for tissue infiltration by innate cells implicated in allergic and helminth immunity and this process can be negatively regulated by a vertebrate chitinase.

Using infection with the migrating helminth, *Nippostrongylus brasiliensis*, to examine lung tissue responses in mice, we confirmed previous findings that genes highly dependent on signal transducer and activator of transcription 6 (Stat6) induced during infection included acidic mammalian chitinase (AMCase) and Ym1 and Ym2 (in humans, these are known as chitinase-3-like proteins 3 and 4, CHI3L3/4, which could not be distinguished on the microarray)⁵. These proteins belong to the nine-membered family of mammalian chitinase-like proteins. Unlike other family members, which were not regulated by Stat6 (Supplementary Fig. 1), we established that AMCase and Ym2 were expressed in the lung as messenger RNA by day 3 after infection and as protein at high levels by day 9 (Fig. 1a). *N. brasiliensis* larvae migrate within hours from the subcutaneous inoculation site to the lung, where organisms moult and ascend the trachea by day 2 and are swallowed to reach the intestines. In mice, adult worms are expelled by day 10 by an immune response dependent upon IL-4 and IL-13 (ref. 6).

The finding that the expression of chitinase-like proteins was Stat6-dependent in the lung raised the possibility that chitin, which constitutes a structural constituent remodelled during moulting in worms⁷, might serve as a recognition element for eliciting tissue infiltration by IL-4- and IL-13-producing immune cells. To test this, we administered chitin to the lungs of 4get mice, which contain a knock-in IL-4 green fluorescent protein (GFP) reporter that allows the detection of cells competent to produce IL-4 (ref. 8). Chitin challenge led

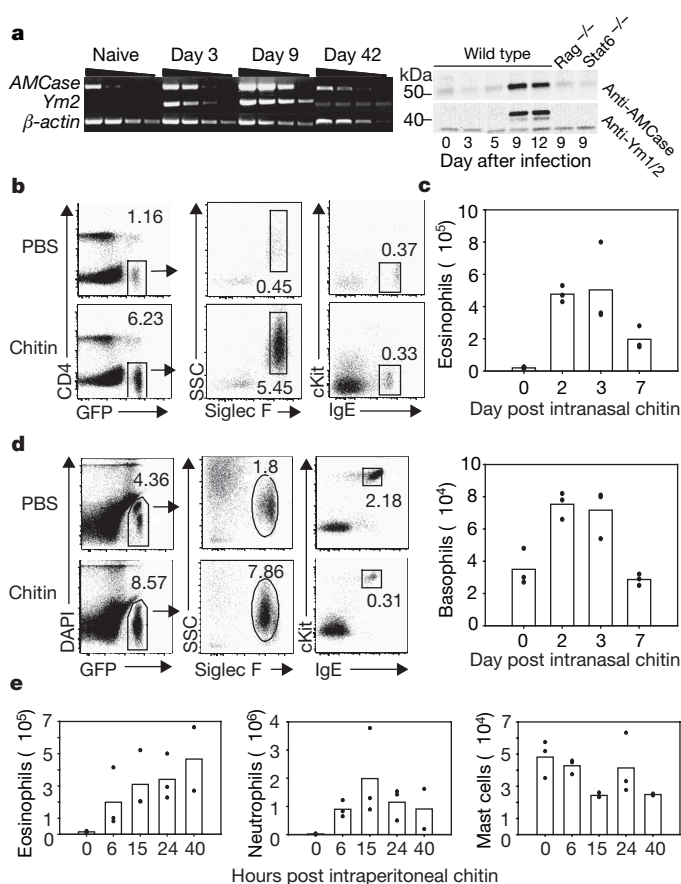


Figure 1 | Chitin induces accumulation of innate effector cells.

a, Polymerase chain reaction with reverse transcription (RT-PCR) and western blot analysis of AMCase and Ym2 in the lung following *N. brasiliensis* infection. **b**, Fluorescence-activated cell sorting (FACS) analysis one day after two intranasal chitin doses to 4get mice. Numbers by gates indicate percentage of live cells. Eosinophil (GFP⁺Siglec F⁺; refs 27, 30) and basophil (GFP⁺, IgE⁺cKit⁺; ref. 27) gates shown. SSC is side-scatter. **c**, Kinetics of chitin-induced lung eosinophils. **d**, FACS of peritoneal cells two days after intraperitoneally injected chitin. Eosinophils and mast cells (GFP⁺IgE⁺cKit⁺) gated. **e**, Kinetics of chitin-induced peritoneal cells. Data representative of two to three independent experiments. In all figures, the dots represent values for individual mice.

¹Howard Hughes Medical Institute, Departments of Medicine and Microbiology/Immunology, University of California San Francisco, San Francisco, California 94143-0795, USA.

²Division of Rheumatology, Allergy and Immunology, Centre for Immunology and Inflammatory Diseases, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts 02129, USA. ³Department of Molecular Cell Biology, Vrije Universiteit, 1091 BT, Amsterdam, The Netherlands. [†]Present address: Institute for Immunology, University of Munich, Munich D-80336, Germany.

to recruitment of eosinophils and basophils to the lungs that peaked by days 2–3 and returned to basal levels by day 9 (Fig. 1b, c). Intraperitoneal injection of chitin induced eosinophil accumulation as early as 6 h later, demonstrating that the response was not tissue-specific (Fig. 1d, e). Neutrophils were also recruited to the peritoneum early in response to chitin, although there was no influx of neutrophils to the lung (Fig. 1e, Supplementary Fig. 2). Unlike in the lung, mast cells (eGFP⁺, cKit⁺, IgE⁺) represent a substantial proportion of constitutively enhanced green fluorescent protein (eGFP)-positive cells in the peritoneum and basophils were not recruited by chitin. Unlike lipopolysaccharide, which resulted in neutrophil recruitment but no eosinophil recruitment to either the lung or peritoneum, chitin induced eosinophil influx in both tissues. Neither challenge altered peritoneal mast cell numbers (Supplementary Fig. 2).

Although lipopolysaccharide can modulate allergic immunity in some models, chitin mediated eosinophil and basophil recruitment in Toll-like-receptor 4 (TLR4)-deficient (lipopolysaccharide-unresponsive) and myeloid differentiation factor 88 (MyD88)-deficient mice (Fig. 2a, and data not shown). Next, we produced recombinant enzymatically active mouse AMCase, which binds and degrades chitin⁹; mutant AMCase with alterations of the active site aspartate 154 and glutamate 158 necessary for the catalytic activity¹⁰; and Ym2, a non-chitinolytic member of the chitinase-like family which is also Stat6-induced but lacks chitin-binding and chitinolytic functions¹¹. Pre-treatment of chitin with enzymatically active AMCase, but not with mutant enzymatically inactive AMCase or with Ym2, led to loss of eosinophil and basophil recruitment to the lung, consistent with a role for intact

chitin in these cellular events (Fig. 2b, c). Similar results were seen after injection of these preparations intraperitoneally to assess eosinophil recruitment (data not shown).

We generated transgenic mice that overexpressed AMCase constitutively in the lung, resulting in the production of enzymatically active AMCase that approximated the levels seen at day 9 after infection with *N. brasiliensis* (Fig. 2d). Analysis of two independent transgenic lines revealed that lung-specific expression of AMCase led to no apparent abnormalities in the mice and the lungs were histologically normal (data not shown). These mice were crossed to 4get mice, thus allowing analysis of the recruitment of IL-4-competent cells to the lung. When challenged with chitin, mice overexpressing AMCase (SPAM×4get) had attenuated inflammatory responses (Fig. 2e). These experiments suggest that chitin itself robustly elicits infiltration by IL-4-competent innate cells following introduction into tissues and that the Stat6-inducible chitinase AMCase abrogates this activity.

Although AMCase expression was Stat6-dependent and recombination activating gene (*Rag*)-dependent, eosinophil recruitment mediated by chitin *in vivo* was independent of Stat6 and *Rag* (Fig. 3a, b), suggesting an early innate response. Basophil recruitment to the lung was also independent of Stat6 and *Rag* (data not shown). Additionally, we could consistently document eosinophil chemoattractant activity in supernatants of bone-marrow-derived macrophages and the RAW267.4 macrophage cell line after incubation with chitin that was unaffected by the addition of neutralizing antibody against eotaxin (data not shown). However, eosinophil

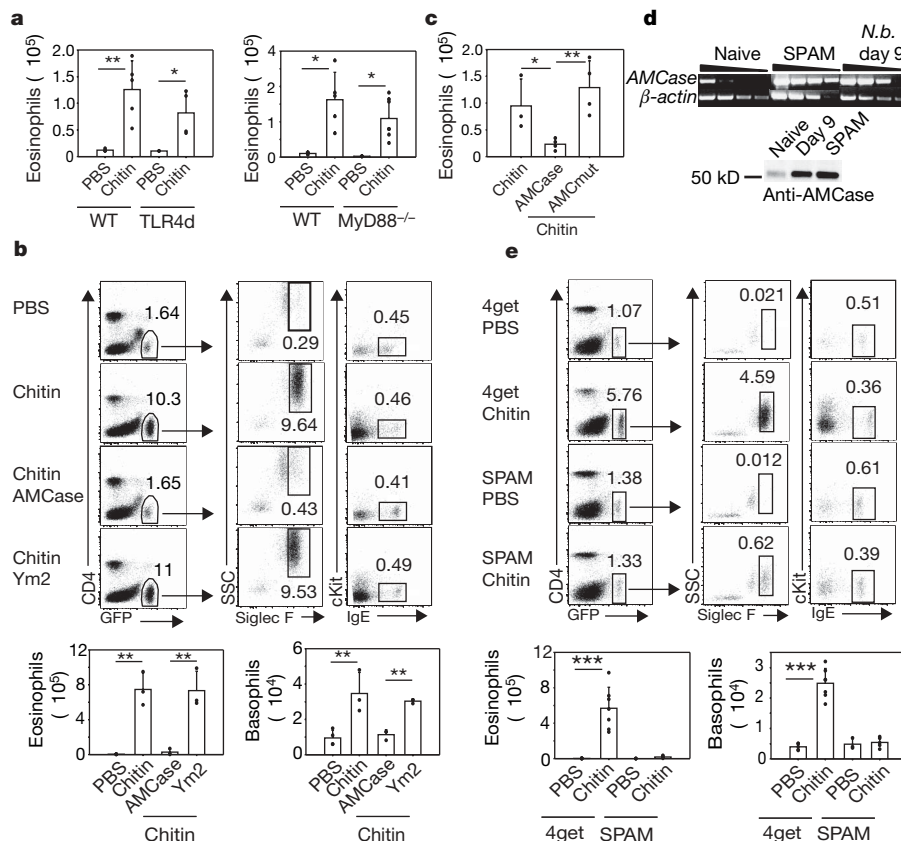


Figure 2 | AMCase prevents chitin-induced eosinophil and basophil recruitment. **a**, Designated mice received intranasal chitin, and lung eosinophils were quantified by flow cytometry. **b**, Chitin or chitin preincubated with recombinant AMCase or Ym2 was administered to 4get mice and recruitment of lung eosinophils and basophils was determined. **c**, Chitin was preincubated with PBS, recombinant AMCase, or recombinant mutant enzymatically inactive AMCase (AMCmut) and given to 4get mice. **d**, AMCase expression in SPAM transgenic mice assessed by RT-PCR and

western blot analysis of lung tissues from naive, day 9 *N. brasiliensis* (*N. b.*)-infected wild-type and naive SPAM mice. **e**, SPAM×4get transgenic mice (SPAM) or wild-type 4get mice were given intranasal chitin. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001 *P*-values were determined using Student's *t*-test. Error bars represent standard deviations. In FACS plots, numbers represent percentage of gated cells from total live cells. Data shown is representative of three independent experiments, with *n* = 3–4 per group.

chemoattraction by macrophages was inhibited if the chitin was pretreated with enzymatically active AMCase, consistent with our *in vivo* results, and was also inhibited when the macrophages were pretreated with an inhibitor of leukotriene production, MK886 (Fig. 3c). Leukotriene B₄ is a potent chemoattractant for eosinophils¹² and is induced from canine phagocytes by chitin¹³. Additionally, the high-affinity receptor for leukotriene B₄, BLT1, is an important mediator of early effector T-cell recruitment to the lung in an allergic asthma model¹⁴.

To establish a role for leukotriene B₄ in the chitin-mediated effects we observed, we challenged mice deficient in BLT1 with chitin. As compared to controls, chitin-induced eosinophil and basophil recruitment to the lung was significantly attenuated in the BLT1 knockout mice (Fig. 3d, and data not shown), despite the equivalent percentages of eosinophils present in the blood of the knockout mice as compared to wild-type mice (Supplementary Fig. 3). To determine

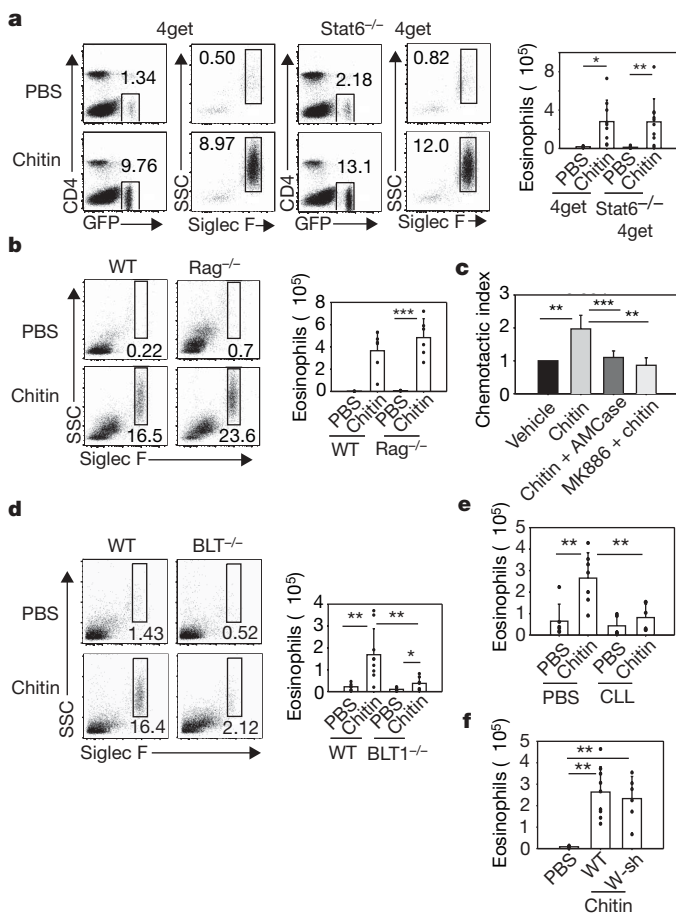


Figure 3 | Chitin-induced eosinophil recruitment is independent of Stat6 and Rag, but is dependent on BLT1. **a**, Stat6^{-/-} × 4get and 4get mice received chitin or PBS, and the lungs were analysed for eosinophils. **b**, Rag^{-/-} and wild-type mice received chitin or PBS, and lungs were analysed for eosinophils. Eosinophils were identified as 4,6-diamidino-2-phenylindole (DAPI)⁻, forward scatter (FSC)^{lo}, CD4⁻, Siglec F⁺, SSC^{hi}. **c**, Transwell analysis of eosinophil recruitment by RAW267.4 macrophages incubated with vehicle, chitin, or AMCase-treated chitin. Where indicated, macrophages were pretreated with vehicle or MK886 (10 μM) for 10 min at 37 °C before addition of chitin. **d**, BLT1^{-/-} mice and wild-type C57BL/6 mice received intranasal chitin. **e**, 4get mice were treated intraperitoneally with clodronate liposomes (CLL) or PBS on days 0 and 3. On day 4 mice received chitin intraperitoneally and were analysed 6 h later. **f**, Mast-cell-deficient (W-sh) or wild-type controls received chitin intranasally. Error bars represent standard deviations. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001 *P*-values were determined using Student's *t*-test. Numbers in FACS plots are percentages of gated cells from total live cells. Experiments represent two to three independent experiments with *n* = 3–4 per group.

whether macrophages play a role *in vivo*, we used clodronate liposomes to deplete peritoneal macrophages before the administration of chitin. Treatment with clodronate markedly curtailed eosinophil recruitment to the peritoneum by chitin (Fig. 3e). Mast cells are also potent producers of leukotrienes. Following administration of chitin to mast-cell-deficient Kit^{W-sh/W-sh} mice¹⁵, eosinophil recruitment to the lungs was unaffected, suggesting that mast cells are not required for a chitin-mediated response (Fig. 3f).

The capacity of chitin to mediate effects on macrophages suggested that these cells might represent sensors for chitin in tissues. A signature gene induced in alternatively activated macrophages is arginase I (*arg1*)^{2,5}. Arginase I enzymatically cleaves arginine to generate L-ornithine, a precursor for polyamines and proline, which have been implicated in cell proliferation and collagen production, respectively. To assess whether chitin induces alternative macrophage activation *in vivo*, we generated mice containing an IRES-driven enhanced yellow fluorescent protein (eYFP) reporter introduced at the 3'-end of the *arginase I* gene (H.-E.L. and R.M.L., manuscript in progress). Under resting conditions, no eYFP-positive macrophages were identified in the lungs or peritoneum of *arg1* reporter mice, designated YARG (Fig. 4a, data not shown). On day 9 after *N. brasiliensis* infection, however, large numbers of eYFP-positive macrophages (CD11b⁺, CD11c⁻, Gr1⁻) were present in both tissues (Fig. 4a, b, data not shown). Thus, parasitic infection elicits accumulation of arginase-I-positive macrophages, in agreement with earlier studies^{2,5}. Next, we injected chitin into YARG mice. As early as 6 h later, and for up to 7 days, arginase-I-positive cells with surface markers consistent with macrophages appeared in the lungs and peritoneum (Fig. 4c, f, data not shown). Further, when sorted into transwell dishes, arginase-I-positive macrophages induced by chitin attracted eosinophils in a manner inhibited by MK886, consistent with a role for these cells in mediating cell recruitment by leukotriene B₄ *in vivo* (Fig. 4d).

Additionally, we crossed YARG and SPAM mice, and challenged the lungs of these mice with chitin. In contrast to YARG mice, which developed alternatively activated macrophages by day 2, YARG × SPAM mice developed greatly attenuated numbers of these cells (Fig. 4e). Similar results occurred if chitin was digested with AMCase before we injected it into animals (Supplementary Fig. 4). Arginase I was induced initially in the peritoneum in macrophages with the characteristic phenotype of resident cells (CD11b^{hi}, F4/80^{hi}, Ly6C⁻, CD62L⁻). Macrophages with the surface phenotype of inflammatory macrophages (CD11b⁺, F4/80⁺, Ly6C⁺, CD62L⁺) did not acquire arginase I expression until later in the response, indicating that resident macrophages respond early to chitin (Fig. 4f).

Whereas previous reports implicated AMCase as a proinflammatory mediator of allergic inflammation⁴, transgenic mice overexpressing AMCase showed no signs of spontaneous inflammation. Rather, our data support a role for AMCase in the feedback attenuation of chitin-induced allergic innate immune responses by enzymatically degrading chitin, thus removing the stimulus for further eosinophil and basophil recruitment. Although inhibition of leukotriene B₄ rendered mice unable to control the helminth *Strongyloides venezuelensis*¹⁶, *Nippostrongylus* was expelled from BLT1-deficient mice (data not shown), suggesting that additional pathways, such as activation of cytokine secretion by helminth proteases¹⁷, may contribute to the immune response against these complex pathogens. Other studies have suggested that chitin may skew immunity away from T-helper-2-mediated allergic responses, although differences in animal models, chitin preparations and delivery, and the potential for contamination by lipopolysaccharide and other microbial activating ligands makes direct comparisons with our studies difficult¹⁸.

Chitin is the second-most abundant biopolymer in nature, with estimates of billions of metric tons produced annually in oceans alone. Despite this prodigious production by phytoplankton and crustaceans, marine sediments contain only trace amounts owing to degradation by chitinolytic marine bacteria¹⁹. Chitin provides

osmotic stability and tensile strength to fungal cell walls³ and scaffolds the rigid exoskeleton in insects²⁰. Nematode chitins are important for eggshell integrity and for structure of the rigid pharynx, including the buccal cavity and grinder, a specialized cuticle that is shed and resynthesized during moulting⁷. The use of chitin by fungi, worms and insects may have driven evolutionary pressure to maintain chitin-recognition molecules in vertebrates akin to those in plants and protochordates^{21,22}. Intriguingly, these data support a role for chitin as a molecular pattern recognized in tissues and linked with the accumulation of innate cells implicated in helminth and allergic immunity, including alternatively activated macrophages, eosinophils and basophils^{2,23}. Those with occupations predicted to have high environmental chitin levels, such as shellfish processors, have a high

incidence of asthma, suggesting that this pathway may play a role in human allergic disease²⁴.

METHODS SUMMARY

Mice. BALB/c mice, C57BL/6 mice, TLR4-deficient mice (C.C3-TLR4^{Lps-d}/J) were purchased from Jackson Laboratories. MyD88-deficient mice²⁵, mast-cell-deficient *Kit^{W-sh/W-sh}* mice¹⁵, IL-4 reporter mice (4get mice)⁸, and BLT1^{-/-} mice have been described²⁶. Stat6^{-/-} mice, Rag^{-/-} mice and IL-5 transgenic mice were backcrossed ten generations to 4get/BALB/c²³. For information on the generation of SPAM and YARG mice see Methods. Mice were maintained in the specific pathogen-free animal facility at the University of California San Francisco, according to institutional guidelines.

Chitin administration. Chitin (New England Biolabs) was washed three times in PBS and large aggregates settled for 2 min. Suspended chitin was collected and diluted 1:4 in PBS. For intranasal administration anaesthetized animals aspirated 50 µl of chitin on consecutive days, and were analysed on day 2. Where designated, 200 µl of chitin was injected once intraperitoneally. In designated experiments, chitin was preincubated with 100 µg recombinant AMCase, Ym2, or mutant AMCase for 1 h at 37 °C. Additional information on recombinant protein production is available in Methods.

Flow cytometry and cell purification. Flow cytometry was performed as described^{23,27}. For additional information and macrophage purification see Methods.

Transwell assay. RAW264.7 macrophage cell line or bone-marrow-derived macrophages were distributed at 1×10^6 cells per well in 24-well plates in serum-free DMEM media. The following morning adherent cells were stimulated with chitin (0.8 µg) (Sigma) or PBS (vehicle control). Optimal chitin concentration was determined separately in a dose response curve. For macrophages sorted from YARG mice, 5×10^5 cells per well were distributed to 24-well plates in RPMI/10% FCS and used immediately in transwell assays. 1×10^6 splenocytes from IL5 transgenic \times 4get mice were placed in the top well of the transwell chamber (Costar, 5.0 µm pore size, Corning) at 37 °C and 5% CO₂ for 3 h. Cells that migrated to the bottom well of the transwell chamber were enumerated using Counting Beads (Caltag) and stained for Siglec-F-PE. MK886 was purchased from Cayman Chemical. The chemotactic index is the number of cells that migrate to the stimulus divided by the number of cells that migrate to vehicle. Experiments with less than 120,000 migrated eosinophils were excluded because of viability concerns.

Immunohistochemistry. Lungs were isolated from YARG mice and processed as described²⁸ using rabbit anti-GFP polyclonal antibody (ab 6556; Novus Biologicals).

In vivo depletion of macrophages. Cl₂MDP (clodronate) was a gift from Roche Diagnostics. Clodronate liposomes were prepared as described²⁹ and 200 µl of liposomes or PBS was injected intraperitoneally on days 0 and 3. Chitin was administered on the next day and peritoneal cell recruitment was analysed 6 h later.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 January; accepted 13 March 2007.

Published online 22 April 2007.

- Ramalingam, T. R., Reiman, R. M. & Wynn, T. A. Exploiting worm and allergy models to understand Th2 cytokine biology. *Curr. Opin. Allergy Clin. Immunol.* **5**, 392–398 (2005).
- Gordon, S. Alternative activation of macrophages. *Nature Rev. Immunol.* **3**, 23–35 (2003).
- Bowman, S. M. & Free, S. J. The structure and synthesis of the fungal cell wall. *Bioessays* **28**, 799–808 (2006).
- Zhu, Z. et al. Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. *Science* **304**, 1678–1682 (2004).
- Nair, M. G., Guild, K. J. & Artis, D. Novel effector molecules in type 2 inflammation: lessons drawn from helminth infection and allergy. *J. Immunol.* **177**, 1393–1399 (2006).
- Finkelman, F. D. et al. Interleukin-4- and interleukin-13-mediated host protection against intestinal nematode parasites. *Immunol. Rev.* **201**, 139–155 (2004).
- Zhang, Y., Foster, J. M., Nelson, L. S., Ma, D. & Carlow, C. K. The chitin synthase genes *chs-1* and *chs-2* are essential for *C. elegans* development and responsible for chitin deposition in the eggshell and pharynx, respectively. *Dev. Biol.* **285**, 330–339 (2005).
- Mohrs, M., Shinkai, K., Mohrs, K. & Locksley, R. M. Analysis of type 2 immunity *in vivo* with a bicistronic IL-4 reporter. *Immunity* **15**, 303–311 (2001).
- Boot, R. G. et al. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *J. Biol. Chem.* **276**, 6770–6778 (2001).
- Boot, R. G., Renkema, G. H., Strijland, A., van Zonneveld, A. J. & Aerts, J. M. Cloning of a cDNA encoding chitotriosidase, a human chitinase produced by macrophages. *J. Biol. Chem.* **270**, 26252–26256 (1995).

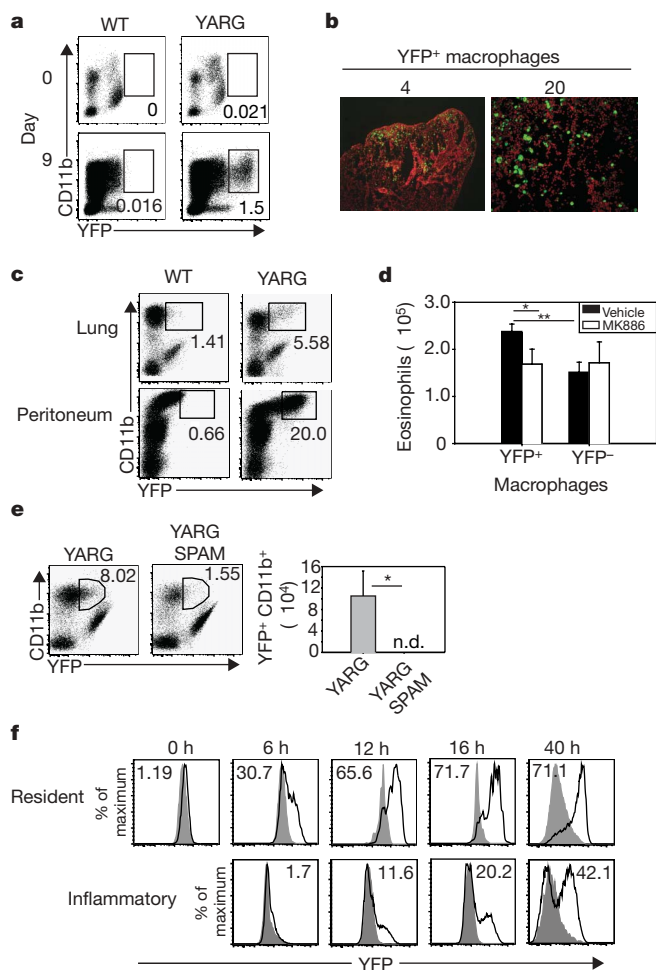


Figure 4 | Chitin induces alternatively activated macrophages. **a**, YARG or wild-type littermates were infected with *N. brasiliensis*. CD11b⁺ lung macrophages were analysed for YFP expression on days 0 and 9. **b**, Day 9 *N. brasiliensis*-infected lung from YARG mice was stained with anti-GFP (green) and nuclei were counterstained with DAPI (red). Lens magnification indicated. **c**, Chitin administered to lung or peritoneum of YARG or wild-type mice. **d**, Sorted YFP⁺ and YFP⁻ macrophages were treated with vehicle or 10 µM MK886 for 10 min at 37 °C before analysis for eosinophil chemotaxis using a transwell assay. Results are an average of three independent experiments. **e**, YARG \times SPAM transgenic mice were compared with YARG littermates following intranasal chitin. **f**, YARG mice or negative littermates received chitin and were analysed for YFP expression. Filled histograms represent negative littermates. The solid line represents YFP in the YARG littermate. Numbers in the histogram are the percentage YFP⁺. Resident and inflammatory macrophages were subset using surface markers described in the text. **P* < 0.05, ***P* < 0.01 *P*-values determined using Student's *t*-test. Error bars represent standard deviations. In FACS plots, numbers represent percentage of gated cell from total live cells. n.d., none detected. Each experiment represents two independent experiments with *n* = 3 per group.

11. Jin, H. M. *et al.* Genetic characterization of the murine *Ym1* gene and identification of a cluster of highly homologous genes. *Genomics* **54**, 316–322 (1998).
12. Huang, W. W. *et al.* Molecular and biological characterization of the murine leukotriene B₄ receptor expressed on eosinophils. *J. Exp. Med.* **188**, 1063–1074 (1998).
13. Usami, Y., Okamoto, Y., Takayama, T., Shigemasa, Y. & Minami, S. Chitin and chitosan stimulate canine polymorphonuclear cells to release leukotriene B₄ and prostaglandin E₂. *J. Biomed. Mater. Res.* **42**, 517–522 (1998).
14. Tager, A. M. *et al.* Leukotriene B₄ receptor BLT1 mediates early effector T cell recruitment. *Nature Immunol.* **4**, 982–990 (2003).
15. Wolters, P. J. *et al.* Tissue-selective mast cell reconstitution and differential lung gene expression in mast cell-deficient Kit^{W^{sh}}/Kit^{W^{sh}} sash mice. *Clin. Exp. Allergy* **35**, 82–88 (2005).
16. Machado, E. R. *et al.* Leukotrienes play a role in the control of parasite burden in murine strongyloidiasis. *J. Immunol.* **175**, 3892–3899 (2005).
17. Phillips, C., Coward, W. R., Pritchard, D. I. & Hewitt, C. R. Basophils express a type 2 cytokine profile on exposure to proteases from helminths and house dust mites. *J. Leukoc. Biol.* **73**, 165–171 (2003).
18. Shibata, Y., Foster, L. A., Bradfield, J. F. & Myrvik, Q. N. Oral administration of chitin down-regulates serum IgE levels and lung eosinophilia in the allergic mouse. *J. Immunol.* **164**, 1314–1321 (2000).
19. Keyhani, N. O. & Roseman, S. Physiological aspects of chitin catabolism in marine bacteria. *Biochim. Biophys. Acta* **1473**, 108–122 (1999).
20. Merzendorfer, H. & Zimoch, L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J. Exp. Biol.* **206**, 4393–4412 (2003).
21. Kaku, H. *et al.* Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proc. Natl Acad. Sci. USA* **103**, 11086–11091 (2006).
22. Hernandez Prada, J. A. *et al.* Ancient evolutionary origin of diversified variable regions demonstrated by crystal structures of an immune-type receptor in amphioxus. *Nature Immunol.* **7**, 875–882 (2006).
23. Voehringer, D., Reese, T. A., Huang, X., Shinkai, K. & Locksley, R. M. Type 2 immunity is controlled by IL-4/IL-13 expression in hematopoietic non-eosinophil cells of the innate immune system. *J. Exp. Med.* **203**, 1435–1446 (2006).
24. Cartier, A. *et al.* Prevalence of crab asthma in crab plant workers in Newfoundland and Labrador. *Int. J. Circumpolar Health* **63** (Suppl. 2), 333–336 (2004).
25. Adachi, O. *et al.* Targeted disruption of the *MyD88* gene results in loss of IL-1- and IL-18-mediated function. *Immunity* **9**, 143–150 (1998).
26. Tager, A. M. *et al.* BLTR mediates leukotriene B₄-induced chemotaxis and adhesion and plays a dominant role in eosinophil accumulation in a murine model of peritonitis. *J. Exp. Med.* **192**, 439–446 (2000).
27. Voehringer, D., Shinkai, K. & Locksley, R. M. Type 2 immunity reflects orchestrated recruitment of cells committed to IL-4 production. *Immunity* **20**, 267–277 (2004).
28. Reinhardt, R. L., Hong, S., Kang, S. J., Wang, Z. E. & Locksley, R. M. Visualization of IL-12/23p40 *in vivo* reveals immunostimulatory dendritic cell migrants that promote Th1 differentiation. *J. Immunol.* **177**, 1618–1627 (2006).
29. Van Rooijen, N. & Sanders, A. Liposome mediated depletion of macrophages: mechanism of action, preparation of liposomes and applications. *J. Immunol. Methods* **174**, 83–93 (1994).
30. Zhang, J. Q., Biedermann, B., Nitschke, L. & Crocker, P. R. The murine inhibitory receptor mSiglec-E is expressed broadly on cells of the innate immune system whereas mSiglec-F is restricted to eosinophils. *Eur. J. Immunol.* **34**, 1175–1184 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Stetson for assistance with recombinant protein production, G. Caughey, X. Xu and S. Akira for mice and reagents, N. Flores, C. McArthur and L. Stowring for expert technical assistance, and A. DeFranco and J. Bluestone for discussions. This work was supported by the NIH, the HHMI and the Sandler Asthma Basic Research Centre at UCSF.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.M.L. (locksley@medicine.ucsf.edu).

METHODS

Mice. For creating SPAM transgenic mice, the full-length complementary DNA for AMCase was isolated from total lung RNA of *N. brasiliensis*-infected BALB/c mice using the following primer pair: 5' AMCase-Sall: 5'-gcgtgcacatggccaagc-tacttc-3', 3' AMCase-BHI: 5'-cgggatccggttcattggccagtg-3'. cDNAs were cloned into pCR2.1 TOPO vector (Invitrogen), sequenced and subcloned into Sall and BglII sites of a lung-specific expression vector containing surfactant protein C promoter elements³¹. The construct was linearized with NotI and injected in DBA/2B6 F1 oocytes. Founder lines with tissue-specific AMCase expression were selected for further study. One line with high lung-specific expression was selected to backcross eight generations onto the BALB/c 4get background.

For generation of YFP-arg1 reporter mice (YARG) a 6-kb BamHI-HindIII fragment was obtained by PCR amplification from 129/SvJ genomic DNA containing exons 4–8 and 3 kb of 3' untranslated sequence of the arg1 gene. BamHI and BglII sites were introduced downstream of the endogenous stop codon and upstream of the 3' untranslated region within exon 8 using PCR-mediated mutagenesis, and the mutated fragment was inserted into pKO-Select-DT (Lexicon) containing diphtheria toxin (DT) α chain for negative selection. A bicistronic reporter cassette⁸ containing IRES linked to eYFP (Clontech) was placed 3' of a floxed-neomycin-resistant cassette derived from pL2neo2. The positive selection/reporter cassette was cloned into the BamHI and BglII sites in the mutated arg1 gene to generate the final targeting construct. PrmCre embryonic stem cells, which express the Cre recombinase under control of the protamine promoter, were electroporated with the NotI-linearized targeting construct and selected in 300 μ g ml⁻¹ G418. Resistant embryonic stem cell clones were screened for homologous recombination by Southern blot. Two independent clones were injected into C57BL/6 blastocysts to generate chimaeras. The floxed neomycin-resistant cassette was deleted in the male germline after breeding male chimaeras to wild-type B6 females. Heterozygous F₁ animals were bred to B6 mice and F₂ offspring were screened for the presence of the reporter and the absence of the PrmCre transgene. YARG mice were backcrossed onto C57BL/6 background for at least six generations.

Flow cytometry and cell purification. Cell suspensions were washed with FACS buffer (PBS, 2% FCS, 1 mg l⁻¹ NaN₃) and the resuspended cell pellets were incubated for 5 min with anti-CD16/CD32 monoclonal antibodies (2.4G2, BD Pharmingen) before staining with APC-Alexa Fluor 750-anti-CD4 (RM4-5, Caltag), PE-anti-Siglec F (E50-2440, BD Pharmingen), APC-anti-ckit (2B8, BD Pharmingen), PerCP cy5.5-anti-CD11b (M1/70, BD Pharmingen), APC-anti-F4/80 (BM8, eBiosciences), and biotinylated anti-IgE (R35-72, BD Pharmingen) followed by PerCP Cy5.5-streptavidin (BD Pharmingen). For some experiments cells were stained with APC-anti-CD62l (BD Pharmingen) or biotinylated anti-Ly6C (BD Pharmingen) followed by APC-streptavidin. Cells were suspended in 1 μ g ml⁻¹ DAPI in FACS buffer to exclude dead cells. Cell counts were determined using Counting Beads (Caltag). For quantification of total YFP⁺ cells in YARG mice the number of background YFP⁺ events detected in wild-type littermates was subtracted from the number of YFP⁺ cells detected in YARG mice. Samples were acquired on a LSRII flow cytometer (Becton Dickinson) and analysed with FlowJo software (Tree Star).

For purification of YFP⁺ and YFP⁻ macrophages, cells were washed from the peritoneum of YARG mice with 10 ml PBS/10% FCS and cell suspensions were stained with PE-anti-CD11b (M1/70, Caltag) and APC-anti-F4/80 (BM8, eBiosciences) in PBS/1% FCS. Cells were suspended in 1 μ g ml⁻¹ DAPI in PBS/1% FCS before sorting (MoFlo; DakoCytomation).

Nippostrongylus brasiliensis infection. Third-stage larvae (L3) of *N. brasiliensis* were recovered from the cultured faeces of infected rats, washed extensively and injected (500 organisms) into mice subcutaneously at the base of the tail. Infected mice were placed on antibiotic-containing water (2 g l⁻¹ neomycin sulphate, 100 mg l⁻¹ chloramphenicol) for five days and killed for analysis after nine days.

Microarray analysis. Total RNA was isolated from the lungs of wild-type and Stat6-deficient mice nine days after *N. brasiliensis* infection using the Total RNA Isolation Kit (Fluka). cDNA was generated using the Superscript reverse transcriptase kit (Invitrogen) and coupled to Cy3 and Cy5 fluorescent dyes (CyScribe dye labelling kit, Amersham Biosciences). Probes were hybridized to spotted glass oligonucleotide (70-mer) arrays, which cover just over 16,400 unique genes (Mouse Genome Set Version 2.0, Qiagen). The arrays were prehybridized in 1% BSA (Invitrogen, Carlsbad, CA), 5 \times SSC, 0.1% SDS for 2 h at 42 °C, washed in water and hybridized with the Cy3/Cy5-labelled probes in 2 \times hybridization buffer (Genesystems) at 53 °C for 45 h. Slides were washed sequentially in 0.03% SDS/1 \times SSC, 0.2 \times SSC and 0.05 \times SSC, dried and scanned on an Axon 4000B scanner. Data were normalized and analysed using Genepix 3.0 software (Axon Instruments, Molecular Devices).

RT-PCR analysis. cDNAs were prepared from total lung RNA generated using the total RNA isolation kit (Fluka) and Superscript reverse transcriptase kit

(Invitrogen). RT-PCR was performed with the following primer pairs: Ym1: 5'-tggaattggtgccctacaa-3' and 5'-aactgcactgtgtatattg-3'; Ym2: 5'-aacgtcagacattcatta-3' and 5'-tggtctctccagtagtaata-3'; Brp39: 5'-agagctgctctgcgtacaag-3' and 5'-agtttctctctgctggctg-3'; AMCase: 5'-tcacaggtctggctctctg and 3'-catatgtcatgacatggatg; β -actin: 5'-atggatgacgatctgct-3' and 5'-atgaggtagtctgcaggt-3'. PCR conditions were 35 cycles with 30 s at 94 °C, 30 s at 58 °C, 60 s at 72 °C followed by a final elongation for 10 min at 72 °C.

Generation of recombinant proteins. Full-length cDNA for Ym2 and AMCase were cloned in pMT/V5-His expression vector (Invitrogen). S2 insect cells were cotransfected with the expression constructs and pCo-Blast selection plasmid conferring blasticidin resistance. After establishing stably transfected cell lines protein production was induced for three days with 0.7 mM CuSO₄ and culture supernatants or control supernatants from untransfected S2 cells were purified over Ni²⁺ columns.

Site-directed mutagenesis of AMCase. Two critical amino acids of the catalytic centre of AMCase, aspartate and glutamate at position 154 and 158, were mutated in AMCase to asparagine and glutamine using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Mutagenesis was performed in two steps using the following primers: aspartate to asparagine 5'-ctggactggcagttaccaggct-cagtgaggagc-3', 5'-gctccacagtgagcctgggtactgccagttccag-3', and glutamate to glutamine 5'-gatggactgaacctggactggcagttaccaggc-3', 5'-gcctgggtactgccagttcaggtt-cagttccatc-3'.

Generation of anti-AMCase antibodies. Antibody was generated using the Invitrogen custom antibody service (Invitrogen). The anti-AMCase antibody was generated against the peptide DKADGLYPVADDRNAFWQC. Briefly, the peptide was conjugated to KLH and the KLH-peptide was injected subcutaneously into three dorsal sites of New Zealand white rabbits with Freund's adjuvant at weeks 0, 2, 7 and 9. Serum was collected at weeks 0, 4, 8 and 10. Terminal bleed of the anti-AMCase antibody was peptide-affinity purified.

Analysis of chitinase activity. Chitinase activity was determined by incubation of recombinant proteins or supernatants from dispersed lung tissue from control or SPAM mice with 270 μ M p-nitrophenyl β -D-N,N',N''-triacylchitotriose (Sigma-Aldrich) in McIlvaine Buffer pH 7.0 (100 mM citric acid, 200 mM sodium phosphate) at 37 °C. *Brugia malayi* chitinase was used as a positive control (New England Biolabs). Substrate turnover by 1 μ g AMCase in 100 μ l volume was analysed on an ELISA reader at 405 nm.

31. Glasser, S. W. *et al.* Genetic element from human surfactant protein SP-C gene confers bronchiolar-alveolar cell specificity in transgenic mice. *Am. J. Physiol.* **261**, L349–L356 (1991).

A positive feedback mechanism governs the polarity and motion of motile cilia

Brian Mitchell¹, Richard Jacobs^{1,2}, Julie Li³, Shu Chien³ & Chris Kintner¹

Ciliated epithelia produce fluid flow in many organ systems, ranging from the respiratory tract where it clears mucus¹ to the ventricles of the brain where it transports cerebrospinal fluid². Human diseases that disable ciliary flow, such as primary ciliary dyskinesia, can compromise organ function or the ability to resist pathogens, resulting in recurring respiratory infections, otitis, hydrocephaly and infertility³. To create a ciliary flow, the cilia within each cell need to be polarized coordinately along the planar axis of the epithelium, but how polarity is established in any ciliated epithelia is not known. Here we analyse the developmental mechanisms that polarize cilia, using the ciliated cells in the developing *Xenopus* larval skin as a model system⁴. We show that cilia acquire polarity through a sequence of events, beginning with a polar bias set by tissue patterning, followed by a refinement phase. Our results indicate that during refinement, fluid flow is both necessary and sufficient in determining cilia polarity. These findings reveal a novel mechanism in which tissue patterning coupled with fluid flow act in a positive feedback loop to direct the planar polarity of cilia.

The skin of amphibian larvae produces a ciliary flow that is invariably oriented along the embryonic axis from anterior-to-posterior (A–P)⁴, a direction that is fixed by patterning of the developing skin during gastrulation^{5–7}. We studied how this A–P polarity is acquired by measuring the polar orientation of cilia using the basal foot, a structure on the basal body at the base of each cilium that points along the polar axis in the direction of the effective stroke (Fig. 1a, c)⁸. By scoring 20 or more basal feet for a given ciliated cell using transmission electron microscopy (TEM) (see Supplementary Fig. S1), we determined both the mean orientation of cilia within that cell relative to A–P embryonic axis (represented by the direction of each arrow in Fig. 1b) and the variation in ciliary orientation around that mean, which is denoted as the circular standard deviation (CSD, represented by the length of each arrow in Fig. 1b, where longer indicates less variation).

Ciliary polarity was first scored at stage 22, when ciliated cells begin to differentiate in the skin but have not yet generated a significant fluid flow. At this stage, cilia roughly point in a posterior direction, but their orientation is imprecise on the basis of two criteria (Fig. 1a, b). First, the mean ciliary direction of the cells varied considerably along the A–P axis (arrow angles in Fig. 1b), indicating that ciliary polarity was not yet precisely coordinated between cells. Second, the CSD of cilia within a given cell also showed considerable variation, with a mean and standard deviation of $53 \pm 13^\circ$. In contrast, by stage 29 when the ciliated cells were producing a strong, directed fluid flow, they have acquired a more defined polarity (Fig. 1c, d). The mean directions of ciliated cells along the A–P axis show less dispersion (compare arrow directions in Figs 1d to b); moreover, the precision of ciliary orientation within individual cells, as reflected by the CSD,

improves to $25 \pm 8^\circ$ (mean \pm s.d., compare arrow lengths in Figs 1d to b). These results indicate that cilia are initially biased in a posterior direction when ciliated cells first differentiate, but then undergo a refinement process in which ciliary orientation along the A–P axis become more precise.

We next determined how cilia orientation is influenced by patterning of the developing skin during gastrulation^{5–7}, by examining cilia polarity after explanting presumptive skin into culture either pre- or post-gastrulation⁹. In control explants made post-gastrulation (stage 16, that is, after patterning), the cilia were as polarized along an A–P axis as in embryos, judged on their ability to produce a directed ciliary flow when tested at stage 29 (Fig. 2a, Supplementary movie 1), as well as the orientation of both the cilia and ciliated cells (Fig. 2b). In contrast, when explants were made pre-gastrulation (stage 10, that is, before patterning), the ciliary flow they produced at stage 29 was chaotic⁹, suggesting alterations in cilia polarity (Figs 2c, Supplementary movie 2). When cilia orientation in these stage 10 explants was assayed at stage 22, the CSD in a given cell

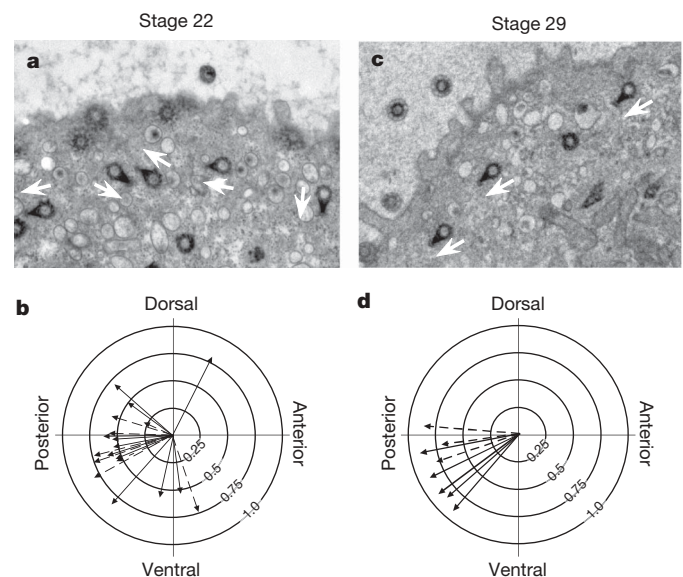


Figure 1 | Analysis of cilia polarity using TEM. Cilia polarity was assayed in embryos by scoring the polar orientation of basal feet using TEM. **a, c**, The polar direction (arrows) of the basal foot was assayed in ciliated cells of stage 22 (**a**) and 29 (**c**) embryos using TEM. **b, d**, Circular plots depicting the polarity of ciliated cells at stage 22 (**b**) and 29 (**d**) in two different embryos (solid and dashed arrows). Each arrow depicts the mean polar direction of cilia in a given cell relative to the embryonic axes, while arrow length represents one minus the circular variance of cilia orientation around the mean.

¹The Salk Institute for Biological Studies, La Jolla, California 92037, USA. ²Howard Hughes Medical Institute, USA. ³Department of Bioengineering and Whitaker Institute of Biomedical Engineering, University of California San Diego, La Jolla, California 92093, USA.

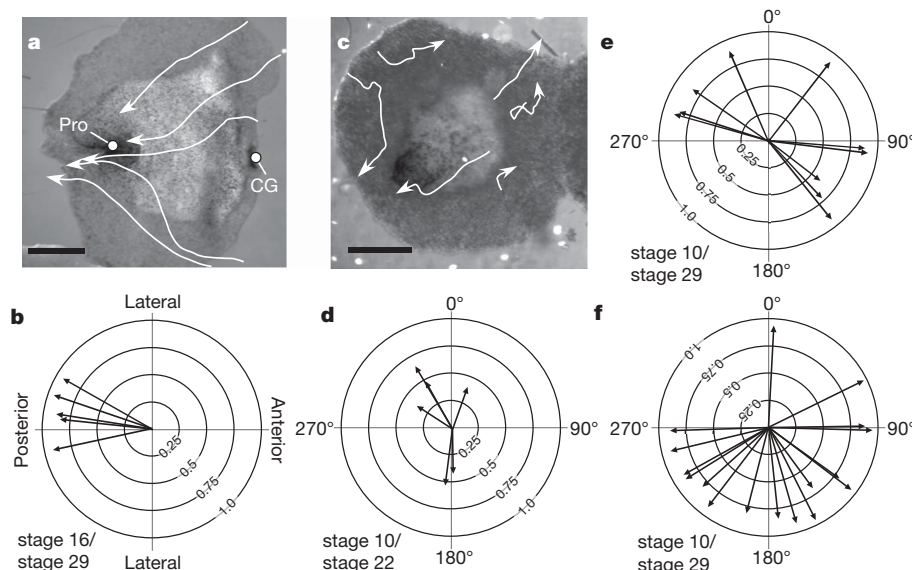


Figure 2 | Mesoderm is critical for polarized fluid flow and cilia direction. **a**, Skin explanted after gastrulation (stage 16) produce a directed fluid flow (white arrows) at stage 29, indicating patterning along an anterior (CG, cement gland) to posterior (Pro, proctodeum) axis. **c**, Skin explanted before gastrulation (stage 10) generates non-directed fluid flow (white arrows) at

stage 29. **b**, **d**–**f**, Circular plots (as in Fig. 1) depicting cilia polarity in stage 16 explants assayed at stage 29 (**b**), or in stage 10 explants assayed at stage 22 (**d**) or at stage 29 (**e**, **f**). The axes in panel **b** are directed according to the A–P axis of the explant, while orientation is arbitrary in panels **c** to **f**. Scale bars in **a** and **b**, 0.5 mm.

($68 \pm 16^\circ$) was only slightly higher ($P = 0.02$) than those in embryos ($53 \pm 13^\circ$). In contrast to embryos, however, the mean ciliary direction of cells in stage 10 explants point randomly around the polar axis (compare arrow directions and lengths in Figs 2d to 1b). At stage 29 after flow begins, the ciliated cells in stage 10 explants still point randomly around the polar axis (compare Figs 2e and f to Figs 2b and 1d), but the variation in planar orientation of cilia within a given cell (CSD = $26 \pm 10^\circ$) is now not significantly different from ($P = 0.9$) that for stage 29 embryos (CSD = $25 \pm 8^\circ$). These results indicate

that each ciliated cell inherently biases the orientation of their cilia along a polar axis, and that patterning of the skin at gastrulation ensures that this bias is directed posteriorly. In the absence of patterning, ciliated cells still refine the orientation of their cilia to polarize in a given direction, but are unable to converge on a common axis.

One common phenotype in human primary ciliary dyskinesia (PCD) is a polar disorientation of cilia, suggesting that cilia polarity is intimately related to cilia function¹⁰. To address this relationship, morpholinos¹¹ were used to knock down the expression of three

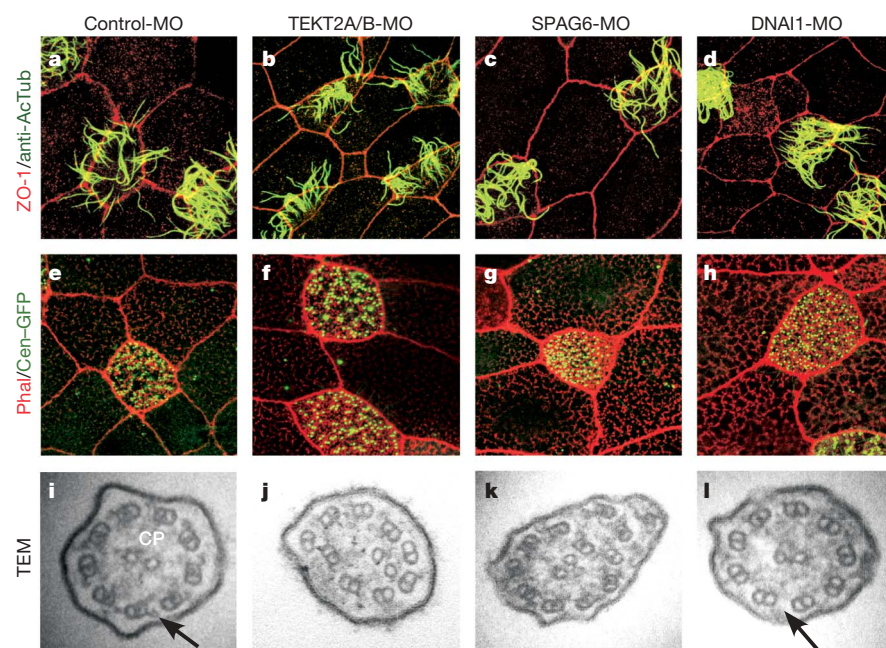


Figure 3 | Morpholinos disrupt cilia structure but not ciliated cell differentiation. Cilia and ciliated cells were imaged in stage 29 tadpoles injected with the indicated morpholino. **a**–**d**, Shown are confocal images of ciliated cells after staining with an antibody to acetylated tubulin (green) that labels cilia, and a ZO-1 antibody (red) that labels cell borders.

e–**h**, Shown are confocal images of ciliated cells after labelling basal bodies with a Centrin2–GFP fusion protein (green) and staining for actin with rhodamine-phalloidin (red). **i**–**l**, Defects in cilia ultrastructure in morphants were visualized using TEM (arrow points to dynein arms in **i** and lack of dynein arms in **l**).

proteins, namely Spag6 (central pair protein), TEK2 (isoform of tektin) and DNAI1 (axonemal dynein, intermediate chain), chosen for their known roles in cilia motility: mutations in DNAI1 account for about 10% of the human cases of PCD³ while knock-out mutations of Spag6 and TEK2 cause a PCD-like disease in the mouse^{12,13}. Injection of a Spag6^{MO}, DNAI1^{MO}, or TEK2A^{MO}/B^{MO} into embryos did not noticeably alter ciliated cell differentiation or ciliogenesis, as determined from staining with an acetylated tubulin antibody (Fig. 3a–d), or basal body localization, as assayed using centrin2–green fluorescent protein (GFP) (Fig. 3e–h). However, each morphant showed blocked or greatly reduced ciliary flow (Figs 4b, Supplementary movies 3 and 4), and immotile or slowed cilia beating (Supplementary movie 5). In addition, each morphant contained defects in the axonemal ultrastructure (Fig. 3i–l) that phenocopied those in mutants, although in the TEK2 morphants these defects were more severe than those reported previously¹³. Targeting these three gene products with morpholinos, therefore, appears to result

in different defects in axonemal structure, thereby disrupting cilia motility and mimicking human PCD.

When analysed by TEM at stage 29, the polarity of cilia in morphant embryos was markedly disoriented in a manner similar to that of unrefined cilia observed in control embryos at stage 22 (Fig. 4c–f). The CSDs of cilia in the morphants ranged from $38 \pm 10^\circ$ to $68 \pm 19^\circ$, thus varying significantly more in polar orientation than in uninjected embryos ($25 \pm 18^\circ$) or those injected with a control morpholino ($24 \pm 12^\circ$) (Supplementary Table 1). The mean directions of ciliated cells in morphants also varied more than those in controls (compare arrow angles in Fig. 4c–f). Finally, like the unrefined ciliated cells in stage 22 embryos, those in morphants were strongly biased in a posterior direction despite the increase in cilia disorientation (Fig. 4c–f). Cilia function, therefore, appears not to be required to bias the polarity of cilia in a posterior direction, but rather during the process of refinement, when their polarity becomes more precise.

Because cilia function is required to produce flow, one possibility is that cilia refine their polar orientation by using the shear stress created by this flow as a cue. Therefore, we tested whether an externally imposed shear stress affects cilia orientation. Prospective skin was explanted before gastrulation (stage 10 explants) onto glass, and then exposed between stages 22 and 28 (about six hours apart) to a constant laminar flow using a flow chamber^{14,15}. The level of shear stress used, 0.5 dyn cm^{-2} , was roughly that generated by the skin at stage 29 and thus considered physiologically relevant (see Methods). As expected, control explants placed in a flow chamber without being subjected to external flow generated non-directed flow lines (similar to Fig. 2c and Supplementary movie 2). In contrast, experimental explants treated with an external flow of 0.5 dyn cm^{-2} responded by generating strongly directed fluid flow aligned with the polarity of the imposed flow (Fig. 5a top, and Supplementary movie 6). Furthermore, while ciliated cells were directed randomly in the untreated explants (similar to Figs 2e, f), those in the flow-treated explants were directed along an axis corresponding to the axis of external flow (Fig. 5a bottom). We note that the mean direction of ciliated cells is still less precise in these flow-treated explants than in normal skin. This could be due to several factors, for example, the difficulty of imposing laminar flow on a tissue that is not completely flat and the difficulty of reorienting cilia that are not already biased in the same direction as the external flow, as normally occurs. Nonetheless, these data provide compelling evidence that fluid flow can be used as a cue to influence planar polarity of cilia.

As a further control, explants were exposed to an external flow that oscillates back and forth with a periodicity of 1 Hz, resulting in no net flow over time. The flow lines generated by the explants treated in this way were random in direction, perhaps even slower and less oriented than in untreated explants, and the mean polar direction of ciliated cells remained random as in untreated controls (Fig. 5b). Thus, cilia, unlike their response to a directed flow, do not change their polarity in response to oscillating flow.

To reorient, motile cilia could actively sense and respond to flow, or alternatively, ciliated cells could sense flow by other means and then reorient their polarity. For example, cells in the kidney are known to sense and respond to flow using primary, non-motile cilia¹⁶. Therefore, we tested whether cilia have to be active not only to generate flow but also to respond to it. As found in embryos, explants generated from embryos injected with Spag6^{MO}, DNAI1^{MO}, or TEK2A^{MO}/B^{MO} produced little or no flow (Fig. 5c top, data not shown) even though ciliated cell differentiation appeared to be normal (data not shown). In addition, as found in the embryos, the orientation of cilia in these explants was unrefined with a CSD ranging from $36 \pm 6^\circ$ in one of the Spag6^{MO} explants to $84 \pm 26^\circ$ in one of the TEK2A^{MO}/B^{MO} explants (Supplementary Table 1). When such morphant explants were exposed to the same externally applied flow that polarized control explants, there was no improvement in the mean direction (Figs 5c–f). These results suggest

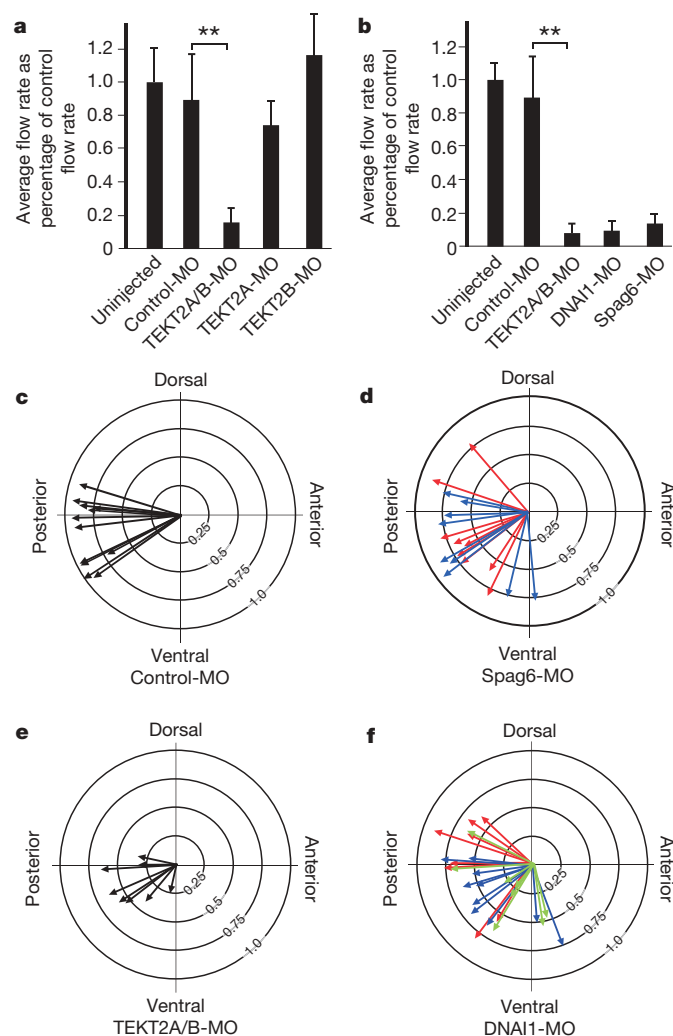


Figure 4 | Cilia polarity is unrefined in tadpoles with immotile cilia. **a, b,** The average rate of fluid flow as a percentage of the control flow rate was determined at stage 29 in tadpoles injected with MOs as indicated. Values shown (mean \pm s.d.) represent the distance travelled per second by a $10 \mu\text{m}$ bead and are based on five measurements from three tadpoles for each condition. Flow values marked with asterisks are significantly different from the control ($P < 0.001$) based on an unpaired, two-tailed Student's *t*-test. **c–f,** Polar plots depicting mean cilia direction and variance (see Fig. 1) of ciliated cells from stage 29 tadpoles injected with a morpholino as indicated. Differently coloured arrows in **d** and **f** depict data collected from different embryos.

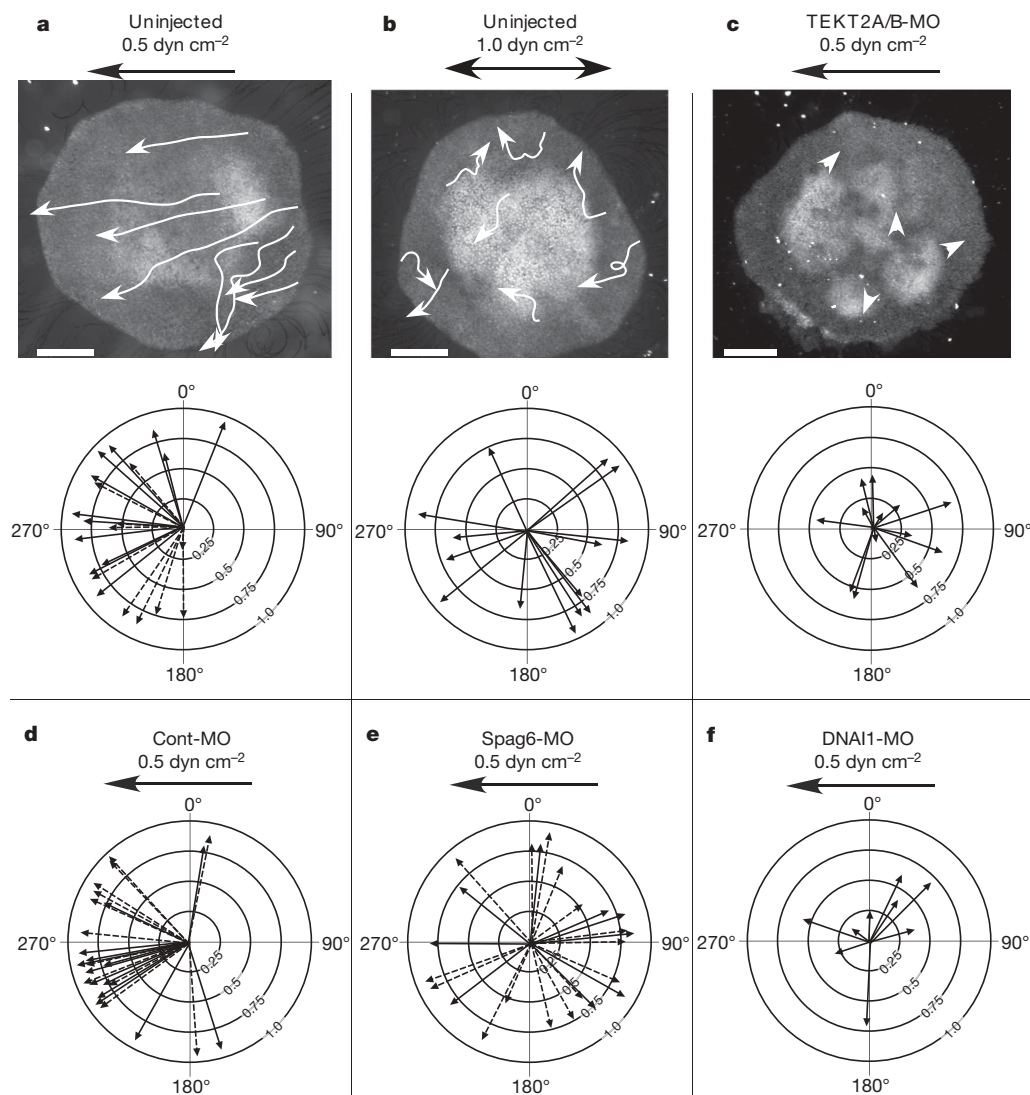


Figure 5 | Cilia polarity can be reoriented by an external fluid flow, but only when cilia are motile. **a–f**, Stage 10 explants were uninjected or injected with the indicated morpholino, cultured to the equivalent of stage 22, placed in a flow chamber, and treated for six hours with either directed, or back-and-forth, fluid flow as shown. After treatment, explants were assayed for fluid

flow by tracking small beads (white arrows in upper panels of **a–c**) and cilia polarity using polar plots (lower panels of **a–c** and panels **d–f**) to depict mean cilia direction and variance (see Fig. 1). Solid and dotted lines in **a, d** and **e** depict data taken from two different explants. Scale bars in upper panels of **a–c**, 0.5 mm.

strongly that motile cilia optimize flow in a positive feedback loop, in which active cilia both generate flow and respond to it by reorientation.

Our findings suggest a stepwise mechanism by which motile cilia acquire polarity in ciliated epithelia. First, an A–P axis is imposed on the presumptive skin during or soon after gastrulation, and ciliated cells use this information to roughly orient their cilia in a posterior direction during their initial differentiation. In the absence of these cues, ciliated cells still choose a polar axis but do so in random directions. Next, this posterior bias allows ciliated cells to produce directional fluid flow, which then acts in a positive feedback loop: as cilia produce flow, they sense it, and modulate their orientation to optimize it. The generation of laminar flow requires a collection of ciliated cells working in unison, so this feedback loop would tend to coordinate polarity over small regions. In addition, feedback is a dynamic mechanism that can respond to changes that occur long after the initial polarity is established, for example, when cells move around as the tadpole grows, or when the skin surface becomes irregular in shape. The mechanism we propose is likely to apply to other epithelia that use ciliary flow to propel fluid or mucus and indeed to cases where ciliary flow fails in human disease¹. For

example, this model accounts for why ciliary disorientation is so commonly associated with ciliary dysfunction in human PCD¹⁷. The positive feedback interplay between fluid flow and orientation of motile cilia of the *Xenopus* larval skin provides new insights into the mechanism for cilia polarization and motion, and this may have broad applications to other forms of motile cilia in the animal kingdom.

METHODS SUMMARY

Embryo injection, explants and flow treatment. *Xenopus* embryos were generated by *in vitro* fertilization and injected with RNAs or morpholinos using standard methods¹⁸. Morpholinos were designed to match the initiation codon of target RNAs (GeneTools)¹¹ and were injected into each quadrant of the animal pole at the two-cell stage using a total of 50–75 ng per embryo. Explants were generated and cultured on fibronectin¹⁹ and treated using a flow chamber^{14,15} with a level of shear stress similar to that produced by a functioning epidermis at stage 29 (see Supplementary Information). Fluid flow produced by larvae or explants was documented by tracking the movement of 10 μ M fluorescent beads using a charge-coupled device (CCD) camera.

TEM and histology. *Xenopus* embryos or explants were fixed, stained either with rhodamine-phalloidin (Invitrogen), or by immunofluorescence using primary antibodies to acetylated tubulin (Sigma) or ZO-1 (Zymed) followed by Cy2- or

Cy3-labelled secondary antibodies, and then imaged using a confocal microscope (Biorad). For TEM, embryos or explants were fixed with 2% glutaraldehyde, postfixed in OsO₄, and stained with uranyl acetate. After embedding in EPON epoxy resin, samples were serially sectioned at 60 nM on an ultramicrotome, collected onto TEM grids while maintaining their orientation, and imaged in the TEM (see Supplementary Fig. 1).

Measurement of cilia orientation. The polar orientation of a ciliated cell was assessed by scoring 20 or more basal feet using AnalySIS image software (Soft Imaging System), analysing these data using Oriana 2.0 (Kovach Computing Services) circular statistics software, and expressing the output on a circular plot as a vector whose angle (μ) denotes the mean ciliary direction and whose length (r) denotes the complement of the circular variance of these cilia around the mean. The average circular standard deviation (referred to as CSD in the text) was calculated using the arrow length for ciliated cells scored under each experimental condition. Values for CSD were compared using a two-tailed, unpaired Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 December 2006; accepted 16 March 2007.

Published online 22 April 2007.

1. Afzelius, B. A. Cilia-related diseases. *J. Pathol.* **204**, 470–477 (2004).
2. Sawamoto, K. *et al.* New neurons follow the flow of cerebrospinal fluid in the adult brain. *Science* **311**, 629–632 (2006).
3. Zariwala, M. A., Knowles, M. R. & Omran, H. Genetic defects in ciliary structure and function. *Annu. Rev. Physiol.* **69**, 423–450 (2006).
4. Assheton, R. Notes on the ciliation of the ectoderm of amphibian embryos. *Q. J. Microscop. Sci.* **38**, 465–484 (1896).
5. Twitty, V. C. Experimental studies on the ciliary action of amphibian embryos. *J. Exp. Zool.* **50**, 310–344 (1928).
6. Woerdeman, M. W. Entwicklungsmechanische Untersuchungen über die Wimperbewegung des Ectoderms von Amphibienlarven. *Roux Arch. Dev. Biol.* **106**, 41–61 (1925).
7. Tung, T. C. & Tung, Y.-F. Y. Experimental studies on the determination of polarity of ciliary action in anuran embryos. *Arch. Biol. (Liege)* **51**, 203–218 (1940).
8. Frisch, D. & Farbman, A. I. Development of order during ciliogenesis. *Anat. Rec.* **162**, 221–232 (1968).
9. König, G. & Hausen, P. Planar polarity in the ciliated epidermis of *Xenopus* embryos. *Dev. Biol.* **160**, 355–368 (1993).
10. Roomans, G. M., Ivanovs, A., Shebani, E. B. & Johannesson, M. Transmission electron microscopy in the diagnosis of primary ciliary dyskinesia. *Ups. J. Med. Sci.* **111**, 155–168 (2006).
11. Nasevicius, A. & Ekker, S. C. Effective targeted gene 'knockdown' in zebrafish. *Nature Genet.* **26**, 216–220 (2000).
12. Sapiro, R. *et al.* Male infertility, impaired sperm motility, and hydrocephalus in mice deficient in sperm-associated antigen 6. *Mol. Cell. Biol.* **22**, 6298–6305 (2002).
13. Tanaka, H. *et al.* Mice deficient in the axonemal protein Tektin-t exhibit male infertility and immotile-cilium syndrome due to impaired inner arm dynein function. *Mol. Cell. Biol.* **24**, 7958–7964 (2004).
14. Frangos, J. A., Eskin, S. G., McIntire, L. V. & Ives, C. L. Flow effects on prostacyclin production by cultured human endothelial cells. *Science* **227**, 1477–1479 (1985).
15. Hsu, P. P. *et al.* Effects of flow patterns on endothelial cell migration into a zone of mechanical denudation. *Biochem. Biophys. Res. Commun.* **285**, 751–759 (2001).
16. Witzgall, R. New developments in the field of cystic kidney diseases. *Curr. Mol. Med.* **5**, 455–465 (2005).
17. Bush, A. *et al.* Primary ciliary dyskinesia: diagnosis and standards of care. *Eur. Respir. J.* **12**, 982–988 (1998).
18. Sive, H., Grainger, R. M. & Harland, R. M. *The Early Development of Xenopus laevis: a Laboratory Manual* (Cold Spring Harbor Press, Plainview, New York, 1998).
19. Davidson, L. A., Hoffstrom, B. G., Keller, R. & DeSimone, D. W. Mesendoderm extension and mantle closure in *Xenopus laevis* gastrulation: combined roles for integrin $\alpha(5)\beta(1)$, fibronectin, and tissue geometry. *Dev. Biol.* **242**, 109–129 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the laboratory for critical comments on the manuscript. We are grateful to C. Stevens and M. Dickinson for discussions. This work was supported by a NIH grant to C.K.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.K. (kintner@salk.edu).

METHODS

Embryos and explants. *Xenopus* embryos were obtained using standard methods¹⁸ and staged according to ref. 20. For explants, the ectoderm (prospective skin) was isolated by excising the animal cap at stage 10, or by removing the skin with or without underlying mesoderm at stage 16. Explants were attached to glass coverslips treated with fibronectin (Roche) and cultured in Danilchik's For Amy (DFA), containing 0.1% BSA as described previously¹⁹.

Morpholinos and injections. Morpholino antisense oligonucleotides (GeneTools) were injected into two-cell stage embryos to inhibit the expression of *Xenopus* *TEKT2*, *Spag6* and *DNAI1*¹¹. In the case of *Spag6* or *DNAI1*, one morpholino was used to target both forms encoded in the *Xenopus laevis* pseudotetraploid genome, while in the case of *TEKT2*, each form was targeted separately using a different morpholino:

TEKT2A (Unigene XL20447): 5'-TCTGAAATGAGGTCACCTTTCTTTAT-3';
TEKT2B (Unigene XL15616): 5'-TCTGTCAGCTACTTCTCAGCCCTGT-3';
Spag6 (Unigene XL46928): 5'-CTTGAAGAACCTGCCGCTGGCTCAT-3';
DNAI1 (Unigene XL16599): 5'-CATCTGTGCTGTTTGTAGTCGGCAT-3';
Control: 5'-GGATGGTGCAGAGTCTCCATCAGTA-3'.

Morpholinos were injected into each quadrant of the animal pole at the two-cell stage using a total of 50–75 ng per embryo.

TEM and histology. Embryos and explants for TEM were fixed overnight at 4 °C in 2% glutaraldehyde, 80 mM cacodylate buffer (CB), pH 7.4, and washed in CB. Post-fixation was performed for 1 h at 4 °C in 1% OsO₄, 1% potassium ferrocyanide and 80 mM CB. Specimens were washed in H₂O and *En Bloc* stained for one hour at 4 °C using 2% uranyl acetate followed by dehydration in increasing amounts of glycol methacrylate. Specimens were embedded in Epon. After embedding, coverglass was dissolved away from explants using concentrated hydrofluoric acid. Serial sections were cut (60 nm thickness) with care taken to maintain proper orientation of the embryo or explant (relative to the direction of flow). Electron micrographs (10,000×) were collected with a digital camera using a Jeol 100 CX II. Images from multiple serial sections were aligned and scored for basal foot orientation.

Ciliated cells were immunostained by fixing embryos or explants with 4.0% paraformaldehyde in phosphate buffered saline (PBS) for one hour on ice. Tissue was stained by overnight incubation in rabbit anti-ZO-1 (Zymed 1:200) and mouse monoclonal anti-acetylated tubulin (Sigma 1:1000) primary antibodies, and a four to six hour incubation in anti-rabbit Cy3 and anti-mouse Cy2 secondary antibodies (Jackson ImmunoResearch). Antibody incubations were performed using PBS containing 0.1% TritonX-100 and 10% heat-inactivated normal goat serum, and washed with PBS containing 0.1% TritonX-100. Basal bodies were labelled by injecting a synthetic messenger RNA encoding centrin2 (UniGene XL50473) fused at the carboxy terminus with GFP. For staining of actin, embryos were incubated in rhodamine-labelled phalloidin (Invitrogen) after fixation.

Circular statistics. Image acquisition and angle determination for basal foot orientation were performed using AnalySIS image software (Soft Imaging System). Statistical analysis and plotting were performed using Oriana 2.0 (Kovach Computing Services) circular statistics software. Each arrow on the polar plot represents a given ciliated cell based on scoring the orientation of at least 20 basal feet in serial sections. The average CSD was calculated by averaging the circular standard deviation for each arrow under each experimental condition. Values for CSD were compared using a two-tailed, unpaired Student's *t*-test.

Flow experiments. A flow chamber was used to impose shear stress as previously described^{14,15}. Tissues explanted onto glass coverslips as described above were placed in a rectangular flow chamber (0.050 cm in height, 2.5 cm in width and 5.0 cm in length) and subjected to a shear stress of 0.5 dyn cm⁻² for six to seven hours during the critical development of stages 22–28. For oscillating flow experiment a shear stress of ± 1 dyn cm⁻² was applied back and forth (no net flow) as a sinusoidal wave at a frequency of 1 Hz. The level of shear stress applied to explants was chosen on the basis of the estimate of the shear stress that is normally generated by flow produced by a functioning epidermis at stage 29. This shear stress was calculated by measuring the velocity of fluid flow occurring at the surface of a stage 29 tadpole by tracking the movement of 10 μ m fluorescent beads (Invitrogen, Molecular Probes) or 0.1% rhodamine dye in 0.1 \times MMR¹⁸, which were placed near the surface by a glass micropipette, and filmed using a digital camera. The average velocity of fluid flow was measured to be 1.7 ± 0.2 mm s⁻¹ and 2.6 ± 0.3 mm s⁻¹ using beads and dye, respectively. The shear stress can be calculated using the following equation²¹: shear stress = (viscosity \times velocity)/distance from the surface. Because the viscosity of 0.1 \times MMR is 0.007 poise and the distance from the surface where the measurements were made was of the order of 30–40 μ m, the calculated shear stress would correspond to approximately 0.4–0.5 dyn cm⁻². Therefore, we chose a value of 0.5 dyn cm⁻² in our flow experiments.

20. Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis* (North Holland, Amsterdam, 1967).
21. Chien, S. in *The Red Blood Cell* (ed. Surgenor, M. D.) 1031–1133 (Academic Press, New York, 1975).

LETTERS

Template switching during break-induced replication

Catherine E. Smith^{1,2}, Bertrand Llorente³ & Lorraine S. Symington¹

DNA double-strand breaks (DSBs) are potentially lethal lesions that arise spontaneously during normal cellular metabolism, as a consequence of environmental genotoxins or radiation, or during programmed recombination processes. Repair of DSBs by homologous recombination generally occurs by gene conversion resulting from transfer of information from an intact donor duplex to both ends of the break site of the broken chromosome¹. In mitotic cells, gene conversion is rarely associated with reciprocal exchange and thus limits loss of heterozygosity for markers downstream of the site of repair and restricts potentially deleterious chromosome rearrangements^{2–5}. DSBs that arise by replication fork collapse or by erosion of uncapped telomeres have only one free end and are thought to repair by strand invasion into a homologous duplex DNA followed by replication to the chromosome end (break-induced replication, BIR)⁶. BIR from one of the two ends of a DSB would result in loss of heterozygosity, suggesting that BIR is suppressed when DSBs have two ends so that repair occurs by the more conservative gene conversion mechanism. Here we show that BIR can occur by several rounds of strand invasion, DNA synthesis and dissociation. We further show that chromosome rearrangements can occur during BIR if dissociation and re-invasion occur within dispersed repeated sequences. This dynamic process could function to promote gene conversion by capture of the displaced invading strand at two-ended DSBs to prevent BIR.

We previously described modification of a chromosome fragmentation assay⁷ as a model system for BIR in *Saccharomyces cerevisiae*⁸. The chromosome fragmentation vector (CFV) contains the *URA3* selectable marker, *SUP11*, *CEN4*, an *ARS* (autonomously replicating sequence), a tract of $(G_1-3T)_n$ to provide a site for telomere addition, and a unique DNA segment from the left arm of chromosome III (containing the *BUD3* gene). After transformation of yeast with the linearized CFV, a chromosome fragment is generated by *de novo* telomere addition at one end of the vector and recombination-dependent replication of 97 kilobases (kb) of chromosomal sequences at the other end⁸. Previous results showed that the efficiency of BIR is only three to fivefold less than the repair of a plasmid-borne DSB by gene conversion^{7–9}. Because cells have the capacity to use either pathway, and the genetic requirements of the two pathways are similar, it is unclear why the repair of two-ended DSBs usually occurs by gene conversion rather than BIR. The choice between these pathways can have profound genetic consequences because BIR between chromosome homologues results in loss of heterozygosity, or in chromosome rearrangements if invasion occurs within dispersed repeated sequences. If one side of a two-ended DSB undergoes strand invasion to prime DNA synthesis, this event must be restricted to prevent extensive replication. The other side of the DSB most probably influences the outcome of the initial strand invasion event so that BIR is permitted from a one-ended DSB, but not from a two-ended DSB¹⁰. The other end of a DSB could engage the donor duplex

by strand invasion or pairing with the displaced strand resulting from the first-strand invasion, and this could potentially restrict DNA synthesis. Alternatively, the initial strand invasion step may result in a short tract of DNA synthesis followed by displacement of the invading strand from the donor duplex, regardless of whether it is a one-ended or two-ended DSB (Fig. 1a). If the other side of the DSB is present and has been resected, annealing could occur between the newly synthesized leading strand and the single-strand tail on the other side of the DSB^{11,12}. However, if the displaced invading strand cannot anneal because the other side of the break is not present, another round of strand invasion would ensue and this cycle would be repeated until homology is found after strand displacement, or until a new telomere is formed when replication reaches the end of the chromosome (Fig. 1a). Studies in *Drosophila* and yeast are consistent with limited homology-dependent DNA synthesis followed by displacement of the invading strand and end-joining repair or single-strand annealing, but no evidence for re-invasion was presented^{13–15}.

To test whether repair can occur by multiple rounds of strand invasion, we used the CFV to transform a diploid strain containing a heterozygous *hphMX4* insertion and restriction site polymorphisms distributed over an 80-kb region of chromosome III (Fig. 1b and 2a)^{16,17}. Chromosome fragments were identified by pulsed field

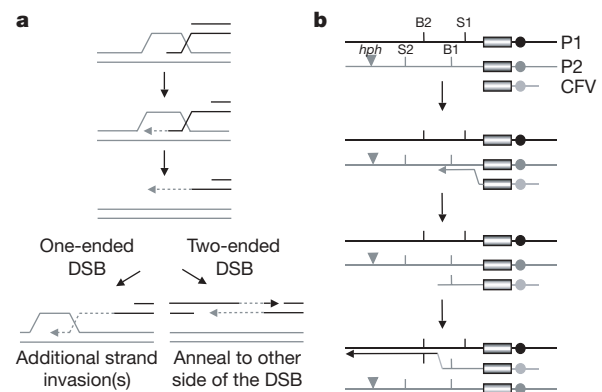


Figure 1 | Multiple strand invasion model and experimental design to test the model. **a**, Following strand invasion, the 3' end is extended by DNA synthesis. The nascent strand is dissociated and can then anneal to the other side of a two-ended DSB. If there is no second end with which to anneal then another round of strand invasion, extension and dissociation occurs. This cycle will repeat until the displaced strand anneals to homologous sequences or forms a new telomere. **b**, A diploid strain containing restriction site polymorphisms downstream of the site of strand invasion was used to test this model. The linear CFV invades one of the two chromosome III homologues; following limited DNA synthesis and dissociation, the second strand invasion could occur into the same or a different—resulting in the formation of a chimeric chromosome fragment—homologue. The shaded box represents the region of homology and the filled circles correspond to centromeres.

¹Department of Microbiology, Columbia University Medical Center, 701 West 168th Street, New York, New York 10032, USA. ²Department of Biological Sciences, Columbia University, 1212 Amsterdam Avenue, New York, New York 10027, USA. ³Unité de Génétique Moléculaire des Levures (URA2171 CNRS and UFR927 Université Pierre et Marie Curie), Institut Pasteur, 25 rue du docteur Roux, F-75724 Paris, Cedex 15, France.

gel electrophoresis (PFGE) of Ura⁺ transformants⁸ and gel slices corresponding to the chromosome III homologues and the chromosome fragments were excised. DNA was extracted from the gel slices and the relevant regions amplified by PCR for restriction analysis (Supplementary Fig. 1)¹⁸. In addition, Southern blot analysis was performed on pulsed-field gels to determine whether the chromosome fragment contained the *hphMX4* insertion (data not shown). Southern blot analysis was also performed on total genomic DNA from chromosome-fragment-containing transformants to eliminate the possibility that chimaeric chromosome fragments were generated by copying of a pre-existing recombinant template formed by a reciprocal exchange between the chromosome III homologues (Supplementary Fig. 1). Of 91 chromosome fragments analysed from wild-type cells, 14 contained sequences derived from both chromosome III homologues and the other 77 had sequences from only one of the two chromosome III homologues (Fig. 2). Of the 14 chimaeric chromosome fragments, two had more than one crossover. For 87 of the chromosome fragment samples analysed the parental chromosomes were unchanged; that is, the DNA recovered from the chromosome III band was heterozygous for all the restriction site markers and in the parental configuration (Supplementary Fig. 1b, lower panel). For two of the 91 samples analysed, one of the chromosome III homologues was not recovered; in both cases the sequences acquired by the chromosome fragment corresponded to the chromosome III homologue that was absent. These rare events most probably occur by a half crossover mechanism. In two

transformants, one of the chromosomal markers had become homozygous.

Analysis of the chromosome fragments recovered using the CFV1 vector showed that most switching occurred within the Bs–Sp interval. To determine whether template switching is restricted to intervals immediately downstream of the site of strand invasion, or is due to an unusual property of the Bs–Sp interval, we constructed a vector (CFV2) that contains sequences homologous to chromosome III 6.6 kb upstream of the region present in CFV1 (Fig. 2). Total genomic DNA was isolated from stable Ura⁺ transformants and analysed by restriction digestion to determine the presence of sites on the chromosome fragments. Eighteen of the seventy events analysed were formed by template switching, and of these two had more than one crossover. After selection for loss of the chromosome fragment on medium containing 5-fluoroorotic acid (selects against Ura⁺ cells)¹⁹, the recombinant fragment was lost in all cases, indicating that it was linked to the chromosome fragment and not due to rearrangement of the chromosome homologues. Similar to the pattern observed with CFV1, most of the switch events were observed in the two intervals immediately adjacent to the site of strand invasion. Only two of the recombinant chromosome fragments generated from CFV2 had a crossover in the Bs–Sp interval—significantly less than observed for CFV1 ($P < 0.005$, Fisher's test), indicating that the Bs–Sp interval is not a recombination hotspot (Fig. 2). These results are consistent with template switching occurring for a limited distance downstream of the site of strand invasion.

If the extended invading strand of the chromosome fragment vector were to dissociate within a dispersed repeated sequence, then we predicted re-invasion might occur at an ectopic repeat, creating a non-reciprocal translocation. Ninety-four independent Ura⁺ transformants derived from CFV1 were analysed by PFGE and Southern blotting to determine whether the chromosome fragments were of uniform size (109 kb). Eleven chromosome fragments were identified with aberrant mobility (Fig. 3); five of these chromosome fragments were analysed by comparative genome hybridization to determine the location of duplicated segments and thus the composition of the chromosome fragment^{20,21}. In all five cases, the vector had invaded at the *BUD3* region of chromosome III as anticipated, copied sequences for about 5 kb downstream of *BUD3* to a δ element (long terminal repeat of a class 2 retrotransposon), and then duplicated sequences downstream of a δ elsewhere in the genome, generating a translocation. These results confirm that template switching occurs by strand invasion, limited DNA synthesis, dissociation of the invading strand and re-invasion.

It is likely that several mechanisms ensure that gene conversion rather than BIR occurs at a two-ended DSB. In a previous study, the kinetics of strand invasion during BIR were shown to be delayed compared with gene conversion, suggesting these two pathways could be differentiated at an early step²². Thus, cells may have a mechanism to sense whether the break has one or two ends and delay strand invasion, or delay DNA synthesis primed from the strand invasion intermediate. Dissociation of the invading strand and second-end capture provides an additional mechanism to ensure that strand invasion intermediates at two-ended breaks are resolved by gene conversion.

We did not detect template switching in 80% of the transformants. Perhaps after the initial strand invasion and displacement the chromosome fragment remains in close proximity to just one homologue and re-invades the same homologue. Alternatively, the chromosome fragments that do not show template switching could have been generated by establishment of a processive mode of replication to complete 97 kb of DNA synthesis from the initial strand invasion event⁶. It is possible that both modes of replication occur in this system, or that an initially unprocessive mode becomes processive as replication proceeds (Fig. 4). All of the template switch events occurred close to the site of strand invasion, consistent with the hypothesis that after several rounds of strand invasion the D-loop

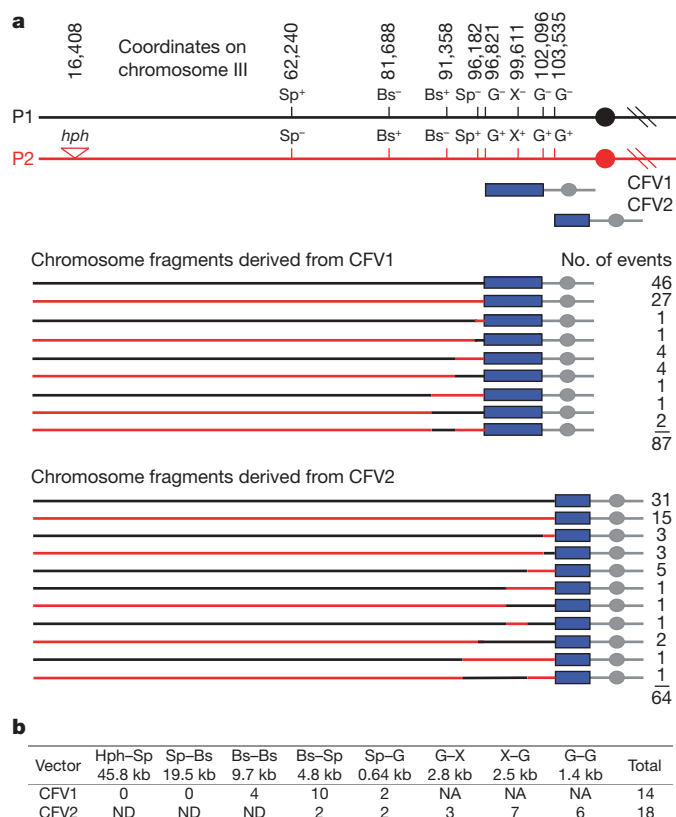


Figure 2 | Chromosome fragments derived from the wild-type diploid show evidence of template switching. **a**, Representation of the chromosome III homologues showing the coordinates from the left end of chromosome III of the heterozygous markers that were scored (not to scale) and the two CFVs used. The lower panels show summaries of the PCR and Southern blot analysis of DNA derived from 87 chromosome fragments from CFV1 and 66 chromosome fragments from CFV2. The chromosome fragments associated with chromosome loss or homozygosity of chromosome III markers are not included in the totals of chromosome fragments presented. **b**, The table shows the total number of switch events for each interval of the two vectors. ND, not determined; NA, not applicable.

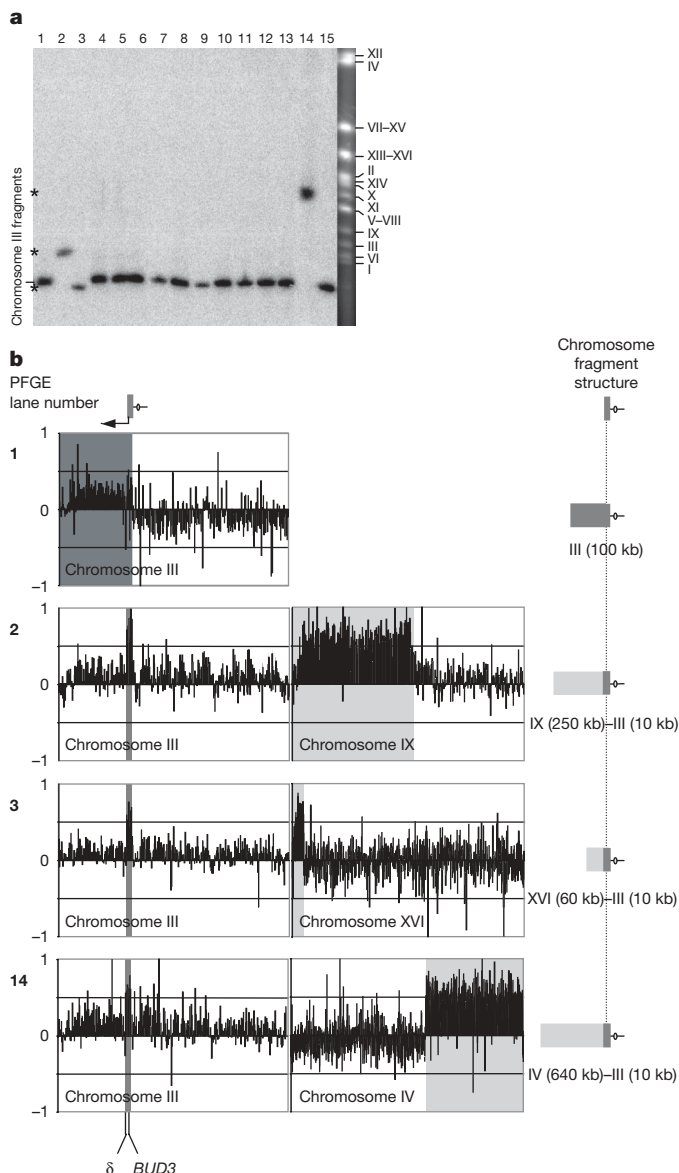


Figure 3 | Formation of non-reciprocal translocations during BIR. **a**, PFGE analysis of 15 transformants; ethidium bromide staining (right lane) and Southern blot analysis using the *URA3* gene as a probe. The asterisks indicate chromosome fragments with abnormal sizes in lanes 2, 3 and 14; chromosome numbers are on the right. **b**, Comparative genome hybridization analysis of the clones from lanes 1, 2, 3 and 14 using the untransformed strain as a reference. The *x* axes represent the ordered yeast open reading frames from left to right for the indicated chromosomes (not to scale). The *y* axes plot the normalized \log_2 ratio of the fluorescence intensities for each open reading frame between the chromosome-fragment-containing strains and the parental diploid strain. The expected \log_2 value for an extra gene copy present on a chromosome fragment is $\log_2(3/2) = 0.58$. Grey zones on the comparative genome hybridization plots represent chromosome regions duplicated by BIR and present on the chromosome fragment. Dark grey represents the chromosome III sequence and light grey is for any other chromosome sequence. Chromosome fragment structures are shown by schematics on the right, and the sizes of the duplicated chromosome regions are indicated. Locations of *BUD3*, where BIR initiates, and the δ element, where template switching occurs, are indicated at the bottom, on chromosome III. Note that *LEU2*, *SUP53* and *Ty2-1* are deleted from the BY4743 diploid; thus, these sequences are not duplicated by chromosome fragment formation.

intermediate is stabilized and processive replication ensues. Cleavage of the crossed strands of the strand invasion intermediate and ligation of the nicked products would be predicted to stabilize the strand

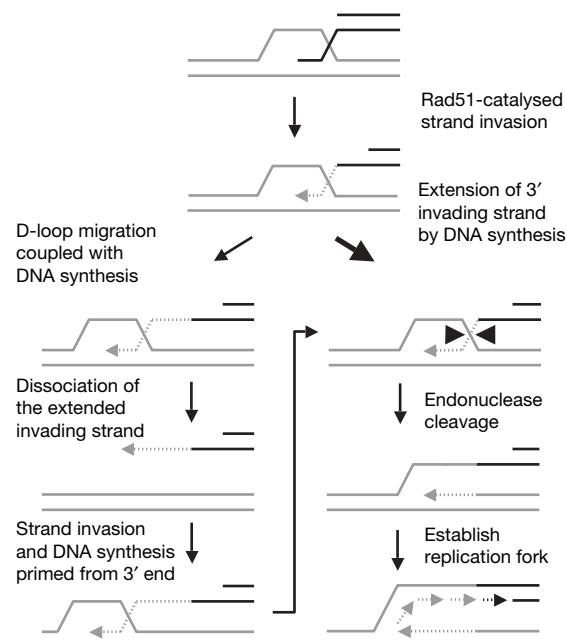


Figure 4 | Models for the post-invasion steps of BIR. The invading strand of the strand invasion intermediate is extended by DNA synthesis forming a migrating D-loop. This structure is unstable and can undergo dissociation and reinvasion. It is unclear when lagging strand synthesis initiates, but it is required to complete replication of the chromosome fragment. Alternatively, the strand invasion intermediate could be cleaved by an endonuclease and the strands ligated to covalently link the invading strand to the donor duplex and establish a stable replication fork.

invasion intermediate and to prevent further dissociation of the invading strand (Fig. 4). Alternatively, template switching may be associated with a region that is temporarily single-stranded by virtue of the delay between initiation of leading strand synthesis primed from the invading strand and initiation of lagging strand synthesis.

The observation that template switching can occur at frequencies as high as 20% during BIR provides strong support for the model of an unstable strand invasion intermediate. We suggest this is a general feature of recombinational repair and not restricted to events that finally resolve as BIR. Studies of gap repair following P-element excision in *Drosophila* are consistent with the requirement for several rounds of strand invasion to complete repair of a large double-stranded gap^{14,15}. Rare triparental recombination events have also been recovered during repair of an HO-endonuclease-induced DSB in mitotic yeast cells²³. An unstable strand invasion intermediate may be a prerequisite for repair by synthesis-dependent strand annealing, but has the potential to generate chromosome rearrangements during BIR. We identified several chromosome fragments with a non-reciprocal translocation between a δ element located 5 kb downstream of the site of strand invasion and a δ element on another chromosome (Fig. 3), providing support for the model that chromosome rearrangements at Ty elements are formed by BIR²¹. The rare and complex chromosome rearrangements detected in *S. cerevisiae* *sgs1 rrm3* double mutants can also be explained by switching between diverged sequence templates during recombination²⁴. The fact that template switching occurs for only a limited distance downstream of the break site may be important in suppressing gross chromosomal rearrangements.

METHODS

Yeast strains and plasmids. The diploid yeast strain with chromosome III markers was made by crossing haploid strains MW16 (*MAT α trp1 arg4 tyr7 ade6 ura3*) and MD249 (*MAT α leu2-bst ade6 ura3 cha1::hphMX6*)^{16,17}. A diploid of the sequenced *S. cerevisiae* strain, BY4743 (*MAT α / α his3 Δ 1/ his3 Δ 1 leu2 Δ 0/ leu2 Δ 0 met15 Δ 0/MET15 LYS2/lys2 Δ 0 ura3 Δ 0/ ura3 Δ 0*) was used to analyse translocations. The chromosome fragmentation vector, CFV/D8B-tg (CFV1),

containing a 5.2-kb insert from the left arm of chromosome III (*Saccharomyces* Genome Database coordinates, 96,821–102,096) was described previously⁸. CFV2 was constructed by replacing the 5.2-kb *Bgl*II fragment of CFV1 with a 3.3-kb *Bgl*II fragment from chromosome III (coordinates, 103,535–106,833).

Yeast transformation and DNA analysis. Transformation of competent yeast cells and analysis of stable Ura⁺ transformants by PFGE was as described previously^{8,25}. DNA analysis to score restriction site markers is described in detail in Supplementary Information. Comparative genomic hybridizations were performed as in ref. 20, using Cy3-labelled DNA (parental strain) and Cy5-labelled DNA (transformants with either the expected or altered size of chromosome fragment); data were normalized using the Goulphar R script²⁶.

Received 19 January; accepted 27 February 2007.

Published online 4 April 2007.

- Symington, L. S. Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. *Microbiol. Mol. Biol. Rev.* **66**, 630–670 (2002).
- Jinks-Robertson, S. & Petes, T. D. Chromosomal translocations generated by high-frequency meiotic recombination between repeated yeast genes. *Genetics* **114**, 731–752 (1986).
- Malkova, A., Ivanov, E. L. & Haber, J. E. Double-strand break repair in the absence of *RAD51* in yeast: a possible role for break-induced DNA replication. *Proc. Natl Acad. Sci. USA* **93**, 7131–7136 (1996).
- Nickoloff, J. A., Sweetser, D. B., Clikeman, J. A., Khalsa, G. J. & Wheeler, S. L. Multiple heterologies increase mitotic double-strand break-induced allelic gene conversion tract lengths in yeast. *Genetics* **153**, 665–679 (1999).
- Stark, J. M. & Jasin, M. Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks. *Mol. Cell. Biol.* **23**, 733–743 (2003).
- McEachern, M. J. & Haber, J. E. Break-induced replication and recombinational telomere elongation in yeast. *Annu. Rev. Biochem.* **75**, 111–135 (2006).
- Morrow, D. M., Connelly, C. & Hieter, P. “Break copy” duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. *Genetics* **147**, 371–382 (1997).
- Davis, A. P. & Symington, L. S. *RAD51*-dependent break-induced replication in yeast. *Mol. Cell. Biol.* **24**, 2344–2351 (2004).
- Bartsch, S., Kang, L. E. & Symington, L. S. *RAD51* is required for the repair of plasmid double-stranded DNA gaps from either plasmid or chromosomal templates. *Mol. Cell. Biol.* **20**, 1194–1205 (2000).
- Aguilera, A. Double-strand break repair: are Rad51/RecA–DNA joints barriers to DNA replication? *Trends Genet.* **17**, 318–321 (2001).
- Ferguson, D. O. & Holloman, W. K. Recombinational repair of gaps in DNA is asymmetric in *Ustilago maydis* and can be explained by a migrating D-loop model. *Proc. Natl Acad. Sci. USA* **93**, 5419–5424 (1996).
- Nassif, N., Penney, J., Pal, S., Engels, W. R. & Gloor, G. B. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell. Biol.* **14**, 1613–1625 (1994).
- Kraus, E., Leung, W. Y. & Haber, J. E. Break-induced replication: a review and an example in budding yeast. *Proc. Natl Acad. Sci. USA* **98**, 8255–8262 (2001).
- Adams, M. D., McVey, M. & Sekelsky, J. J. *Drosophila* BLM in double-strand break repair by synthesis-dependent strand annealing. *Science* **299**, 265–267 (2003).
- McVey, M., Adams, M., Staeva-Vieira, E. & Sekelsky, J. J. Evidence for multiple cycles of strand invasion during repair of double-strand gaps in *Drosophila*. *Genetics* **167**, 699–705 (2004).
- Merker, J. D., Dominska, M. & Petes, T. D. Patterns of heteroduplex formation associated with the initiation of meiotic recombination in the yeast *Saccharomyces cerevisiae*. *Genetics* **165**, 47–63 (2003).
- White, M. A. & Petes, T. D. Analysis of meiotic recombination events near a recombination hotspot in the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* **26**, 21–30 (1994).
- Dunham, M. J. *et al.* Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **99**, 16144–16149 (2002).
- Boeke, J. D., Lacroute, F. & Fink, G. R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197**, 345–346 (1984).
- Kozul, R., Caburet, S., Dujon, B. & Fischer, G. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23**, 234–243 (2004).
- Lemoine, F. J., Degtyareva, N. P., Lobachev, K. & Petes, T. D. Chromosomal translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites. *Cell* **120**, 587–598 (2005).
- Malkova, A., Naylor, M. L., Yamaguchi, M., Ira, G. & Haber, J. E. *RAD51*-dependent break-induced replication differs in kinetics and checkpoint responses from *RAD51*-mediated gene conversion. *Mol. Cell. Biol.* **25**, 933–944 (2005).
- Ray, A., Machin, N. & Stahl, F. W. A. DNA double chain break stimulates triparental recombination in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **86**, 6225–6229 (1989).
- Schmidt, K. H., Wu, J. & Kolodner, R. D. Control of translocations between highly diverged genes by Sgs1, the *Saccharomyces cerevisiae* homolog of the Bloom's syndrome protein. *Mol. Cell. Biol.* **26**, 5406–5420 (2006).
- Schwartz, D. C. & Cantor, C. R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75 (1984).
- Lemoine, S., Combes, F., Servant, N. & Le Crom, S. Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics* **7**, 467 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. K. Holloman and members of the Symington and Dujon laboratories for comments on the manuscript, and T. Petes for providing the haploid strains with restriction site polymorphisms. These studies were supported by grants from the NIH, and by a postdoctoral fellowship from La Ligue Contre Le Cancer to B.L.

Author Contributions C.E.S. characterized template switching between CFV1 or CFV2 and the marked chromosome III homologues. B.L. designed and performed the experiments and analysed the data characterizing chromosome translocations by comparative genome hybridization. L.S.S. designed the experiments, analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.S.S. (lss5@columbia.edu).

LETTERS

A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein

K. Hun Mok^{1,2}, Lars T. Kuhn¹, Martin Goez³, Iain J. Day^{4,†}, Jasper C. Lin⁵, Niels H. Andersen⁵ & P. J. Hore¹

Insights into the conformational passage of a polypeptide chain across its free energy landscape have come from the judicious combination of experimental studies and computer simulations^{1,2}. Even though some unfolded and partially folded proteins are now known to possess biological function³ or to be involved in aggregation phenomena associated with disease states^{1,4}, experimentally derived atomic-level information on these structures remains sparse as a result of conformational heterogeneity and dynamics. Here we present a technique that can provide such information. Using a 'Trp-cage' miniprotein known as TC5b (ref. 5), we report photochemically induced dynamic nuclear polarization NMR⁶ pulse-labelling experiments that involve rapid *in situ* protein refolding^{7,8}. These experiments allow dipolar cross-relaxation with hyperpolarized aromatic side chain nuclei in the unfolded state to be identified and quantified in the resulting folded-state spectrum. We find that there is residual structure due to hydrophobic collapse in the unfolded state of this small protein, with strong inter-residue contacts between side chains that are relatively distant from one another in the native state. Prior structuring, even with the formation of non-native rather than native contacts, may be a feature associated with fast folding events in proteins.

Experimental advances in nuclear magnetic resonance (NMR) spectroscopy have led to the characterization of a diverse range of unfolded states of proteins⁹. In many cases the presence of residual structure has been shown^{10–13}, but with some significant exceptions¹⁴ the poorly resolved spectra of the unfolded state, arising from conformational exchange and dynamic averaging, have generally hampered structural analysis by NMR. We report here the use of an NMR technique that circumvents some of these problems by transferring information from the unfolded or partially folded states of proteins to the native state¹⁵. Using this 'pulse-labelling' strategy, we probe the unfolded state of TC5b (relative molecular mass $M_r = 2,169$), a miniprotein that possesses key structural hallmarks normally only found in larger proteins⁵. Despite its high relative contact order, TC5b is an ultrafast folding protein¹⁶. This property, together with its small size, has made it an attractive subject for molecular dynamics folding simulations^{17–21}.

Three methodologies are combined in this 'pulse-labelling' experiment (Fig. 1). (1) Photo-CIDNP (chemically induced dynamic nuclear polarization)^{6,22}, a technique for enhancing the NMR signals ('hyperpolarization') of solvent-accessible tryptophan, tyrosine and histidine side chains by means of a laser-induced reaction of the protein with a flavin photosensitizer. (2) Rapid homogeneous mixing of solutions in the NMR sample tube to trigger the folding of a denatured protein on a timescale faster than nuclear spin–lattice

relaxation (Supplementary Fig. 1)^{7,8}. To these two techniques we add here, for the first time, (3) transfer of nuclear magnetization via nuclear Overhauser effects (NOEs) from the hyperpolarized side chain protons to neighbouring atoms before the refolding step. As a result, inter-residue contacts in unfolded conformations can be detected in the well-resolved NMR spectrum of the refolded native state.

¹H photo-CIDNP measurements were initially performed on the native and denatured states of TC5b. The photo-CIDNP spectrum of native TC5b (Fig. 2b) is considerably simpler than the conventional NMR spectrum (Fig. 2a), because only two of the 20 side chains, Tyr 3

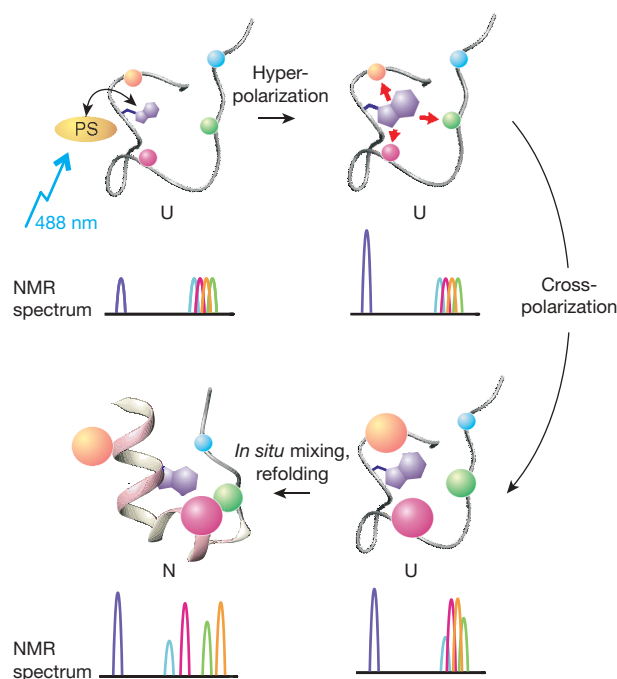


Figure 1 | The photo-CIDNP pulse labelling technique. Laser excitation of a freely-diffusing photosensitizing molecule and its subsequent interaction with a solvent-accessible aromatic amino-acid side chain (violet) results in a non-Boltzmann nuclear spin state, which further transfers its polarization via NOEs to neighbouring side chains (coloured spheres). For unfolded or partially folded proteins, detection is carried out after refolding the protein to its native state, where the greater chemical shift dispersion in the NMR spectrum allows identification of individual atoms. The increase in the size of spheres represents the extent of cross-polarization. PS, photosensitizer; U, unfolded state; N, native (folded) state.

¹Department of Chemistry, University of Oxford, Physical & Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, UK. ²School of Biochemistry and Immunology, Trinity College, University of Dublin, Dublin 2, Ireland. ³Fachbereich Chemie, Martin-Luther-Universität Halle-Wittenberg, Kurt-Mothes-Strasse 2, D-06120 Halle/Saale, Germany. ⁴Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605, USA. ⁵Department of Chemistry, University of Washington, Seattle, Washington 98195, USA. [†]Present address: Medway Sciences, University of Greenwich, Medway University Campus, Central Avenue, Chatham Maritime, Chatham, Kent, ME4 4TB, UK.

and Trp 6, are hyperpolarizable. The enhancements of the two aromatic side chains reflect the rates of their reaction with the photo-excited flavin^{6,22}.

Figure 2b–d shows a series of photo-CIDNP spectra of native TC5b recorded with increasing laser irradiation times but keeping the delivered energy constant per flash. While the signal intensities from the directly polarized protons remain approximately constant, there are resonances in the aliphatic region (principally between 0.5 and 2.1 p.p.m.) that are essentially absent in Fig. 2b but which grow in intensity as the irradiation time is increased. These are NOE peaks arising from transfer of polarization from Tyr 3 and Trp 6 via dipolar cross-relaxation²³; they correlate well with the cross-peaks between protons separated by 2.0–3.5 Å previously observed in two-dimensional nuclear Overhauser effect spectroscopy (NOESY) spectra of the native state⁵ (Supplementary Fig. 2). The mixture of absorptive and emissive signals in the aliphatic regions of Fig. 2c and d reflect the different phases of the hyperpolarized tyrosine and tryptophan nuclei²³.

In contrast, the NMR (Fig. 2e) and photo-CIDNP spectra of TC5b unfolded in 6.0 M urea (Fig. 2f–h) differ noticeably from those of the native state. The polarized aromatic Trp 6 signals are stronger than those of Tyr 3; together with their linewidth differences and chemical shift movements, this suggests a less structured and more accessible environment for Tyr 3 in the denatured state. The NOESY spectra of the denatured state (Supplementary Fig. 2) show that almost all of the peaks have different chemical shifts. Furthermore, most of the amide and aromatic protons are not well-resolved in the spectrum of the denatured state, making direct resonance assignment extremely difficult. Nonetheless, the photo-CIDNP spectra at longer irradiation times (Fig. 2f–h) do display NOEs.

The pulse-labelling experiment (Fig. 1) entails production of photo-CIDNP and build-up of NOEs in the denatured state, followed

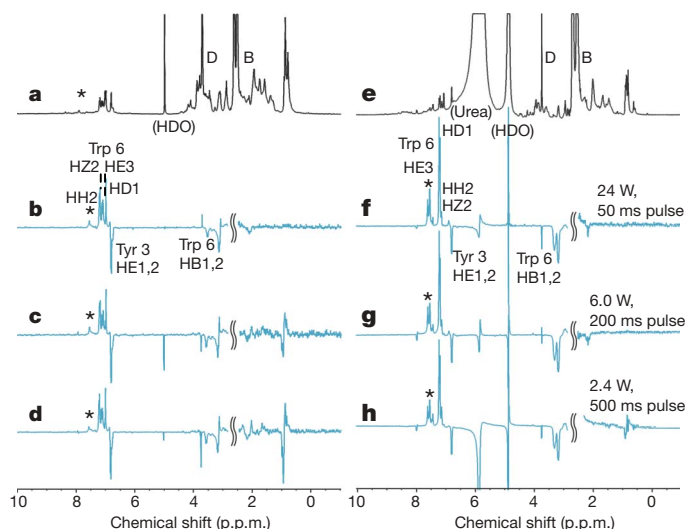


Figure 2 | The 600 MHz ^1H NMR and photo-CIDNP spectra of TC5b. a–d, In its native state; **e–h,** in its denatured state. **a, e,** ^1H NMR spectra for the native state (**a**) and the 6.0 M urea-denatured state (**e**) of TC5b at 5 °C. For photo-CIDNP spectra **b–d** and **f–h**, the power and duration of the laser irradiation were varied as shown. The aliphatic region is magnified four times to aid visualization. Chemical shift assignments of the hyperpolarized side chains of Tyr 3 and Trp 6 are given, together with those of water (HDO), the chemical shift reference (1,4-dioxane, D), the buffer (citrate, B) and urea. In the native state spectra **c** and **d**, the strong emissive cross-polarized peaks of Ile 4 HG2, HD1 and Leu 7 HD1 (~0.9 p.p.m.) arise from the proximity of the Tyr 3 HE2 protons in the native state. Likewise, the absorptive signals from Pro 12 HG (~2.08 p.p.m.) and Leu 7 HG (1.6 p.p.m.) reflect NOEs from Trp 6. Additional small signals (labelled with asterisks) arise from minor conformer populations associated with *cis* isomers of the four proline residues, which all have *trans* conformations in the completely folded native state and in any conformations that can rapidly achieve the folded state.

by rapid injection of the protein solution into refolding buffer in the NMR sample tube. A free induction decay is acquired after a specified delay (100 ms) to allow complete mixing *in situ*. The resulting CIDNP spectrum of the refolded native state exhibits aliphatic resonances whose amplitudes reflect the CIDNP-NOEs that were generated before folding. TC5b is well suited to this procedure, as refolding has been shown to be very rapid (the time constant for refolding being 4 μs ; ref. 16).

The pulse-labelled ('Unfolded-to-Native', U-to-N) spectrum of TC5b (Fig. 3b) differs considerably from the conventional CIDNP spectrum of native TC5b (Figs 2d and 3d), especially with regard to the prominent signals in the aliphatic region (0.5–2.0 p.p.m.). Control experiments in which conditions were chosen so that the injection resulted in no change of state of the protein were also performed: U-to-U and N-to-N (Fig. 3a and c, respectively). These control spectra strongly resemble the corresponding conventional CIDNP spectra (Fig. 2h and d, respectively), supporting the conclusion that the aliphatic signals in the U-to-N pulse-labelled spectrum truly reflect inter-residue contacts in the unfolded ensemble.

The U-to-N NOE peaks (Fig. 3b) can be identified straightforwardly from the NMR chemical shift assignments of the native state (Supplementary Fig. 2 and Supplementary Table 1). Four side chains were found to be recipients of cross-polarization: Ile 4, Leu 7, Pro 12 and Arg 16. In addition, because Tyr 3 is largely exposed to the solvent and well-removed from neighbouring side chains (Supplementary Fig. 3), Trp 6 is likely to be the major source of the NOEs seen in the U-to-N spectrum. Average distances of the Ile 4, Leu 7, Pro 12 and Arg 16 protons from the Trp 6 side chain in the unfolded state can be estimated as detailed in the Supplementary Methods. Using experimentally determined ^1H spin-lattice relaxation times for native and unfolded states (Supplementary Table 2), estimates of the proton-proton distances are obtained using, as a calibration, the ensemble-averaged distance of 3.41 Å from Trp 6 HD1 to the Arg 16 HB2 proton in the TC5b native state (Table 1).

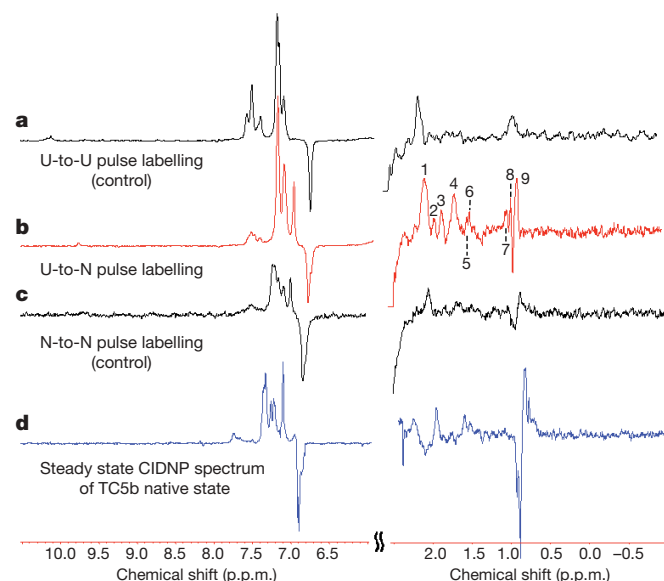


Figure 3 | The 600 MHz photo-CIDNP pulse-labelled NMR spectra of TC5b. The laser irradiation time was 500 ms at a power of 5.0 W. **a,** Pulse-labelled control spectrum of the unfolded state of TC5b in 6.0 M urea ('U-to-U pulse labelling'). **b,** Pulse-labelled spectrum of the unfolded state of TC5b ('U-to-N pulse labelling'). **c,** Pulse-labelled control spectrum of the native state of TC5b in 0.55 M urea ('N-to-N pulse labelling'). **d,** Steady-state CIDNP spectrum of the native state of TC5b (identical to Fig. 2d). The assignment of peaks in the aliphatic region is as follows: 1, Pro 12 HB2 and HG2; 2, Ile 4 HB; 3, Arg 16 HB2; 4, Arg 16 HB3; 5, Leu 7 HG; 6, Ile 4 HG2; 7, Leu 7 HD1; 8, Ile 4 HG2, HD1; and 9, Leu 7 HD2.

Table 1 | Estimated distances of cross-polarized nuclei from Trp 6

	Cross-polarized protons					
	Ile 4 HB	Ile 4 HG2 HD1	Leu 7 HD1	Leu 7 HD2	Pro 12 HG2	Arg 16 HB2 HB3
Unfolded state (Å)	4.2	4.5	4.0	3.4	3.3	4.2
Native state (Å)*	8.07†	8.71† 8.75†	4.81†	3.73†	4.02‡	3.24§ 4.68§

Distances observed by pulse-labelled photo-CIDNP of TC5b. The nearest hyperpolarized Trp 6 proton: †HE3, ‡HH2, §HE1.

* The data for the native state are ensemble-averaged distances to the nearest hyperpolarized aromatic proton of Trp 6 in the NMR structure (ref. 5).

The NOE contact distances observed between Trp 6 and Ile 4, Leu 7, Pro 12 and Arg 16 sites in the unfolded state are, with one exception, shorter than those in the compact native state (Table 1). Although the NOEs in the U-to-N spectra could be attributed to independent close contacts in different conformers rather than a more compact ensemble characterized by a hydrophobic collapse, this would require unrealistically short distances in individual conformers, particularly for the Trp 6/Pro 12 interaction. The hydrodynamic radius of the unfolded state, obtained from diffusion NMR experiments (8.0 ± 0.2 Å; Supplementary Fig. 4), is slightly larger than that of the native state (7.4 ± 0.1 Å), but not nearly large enough to be consistent with a fully unfolded, random coil structure²⁴. Thus, the short distances observed appear to result from simultaneously co-existing contacts within a compact ensemble of structures. Figure 4a shows a representation of a partially collapsed state consistent with the U-to-N NOE observations. A notable non-native interaction involves Ile 4, which lies on the opposite side of the amino-terminal helix and hence is not within NOE range in the native state (Fig. 4b). In addition, although Arg 16 forms a stabilizing salt bridge with Asp 9 in the native state⁵, it is the aliphatic part of the Arg 16 side chain that is found to be associated with Trp 6 in the unfolded state. Furthermore, native-state NOEs between Trp 6 and Pro 17–19 are absent in the U-to-N spectrum. Instead, and in agreement with chemical shift melting data⁵, the unfolded state has a closer association of Trp 6 and Pro 12. Our conclusion regarding a pre-formed hydrophobic cluster in TC5b is consistent with spectroscopic results that show that the denatured state ensemble of the Ile4Gly substitution mutant contains a larger fraction of random coil structures²⁵, and that Trp 6 is situated within a hydrophobic environment in the thermally-denatured state²⁶.

What implications does the existence of structural features in the unfolded ensemble have for the fast folding of TC5b? The existence,

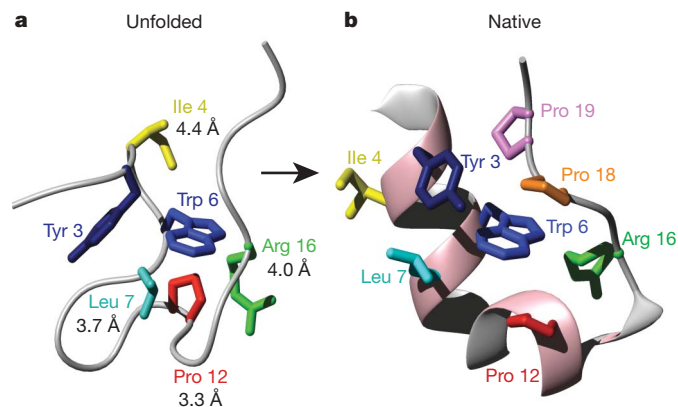


Figure 4 | Representation of the three-dimensional structures of TC5b. **a**, Unfolded state; **b**, native state. Cross-polarized side chains identified in the pulse-labelled spectrum (Ile 4, Leu 7, Pro 12, Arg 16) are indicated in **a**, and the approximate mean distances from the cross-polarization source (Trp 6) are shown. The ribbon structure of the denatured state in **a** is purely schematic; however, structural features are generally drawn according to experimental results given in this and other works (see text for citations).

in simulated folding trajectories, of the Asp 9/Arg 16 salt bridging before achieving the native orientation of the Trp 6 side chain with respect to the prolines has been viewed either as a potential kinetic trap²¹ or as an accelerating feature¹⁸. In the case of a Pro12Trp variant designed to favour prior formation of an edge-to-face Trp 6/Trp 12 interaction, Trp-cage folding is further accelerated²⁷, suggesting that the close interaction between Trp 6 and Pro 12 found here may contribute to the fast folding of the protein. In another case, fluorescence studies on TC5b analogues have shown that a decrease in folding speed is found when elements of the residual structure in the denatured state are removed²⁵. Simulation studies with an extended starting structure suggest that there is an initial collapse event, with different hydrophobic interactions formed depending on the specific trajectory^{17,18}; these interactions may determine the further course of the simulated folding. In agreement with some of the interactions found *in silico*¹⁸, we find that both native and non-native specific contacts pre-exist in the unfolded ensemble. Hence, it appears that the pre-formation of hydrophobically collapsed structures in the equilibrated unfolded state may be a significant factor in accelerating folding²⁸.

The CIDNP pulse-labelling technique can be applied generally to other proteins, provided that a suitable aromatic side chain is present and refolding is fast enough to compete with the relaxation of the nuclear hyperpolarization. Applications to proteins that fold more slowly than ~ 100 ms will require either more dramatic hyperpolarizations^{22,29} or slower nuclear spin-lattice relaxation³⁰. We therefore anticipate that this approach will provide valuable information on the structures of denatured states of many proteins, particularly those that refold rapidly.

METHODS

In situ rapid mixing injection device. The *in situ* rapid mixing injection device recently developed and used for real-time protein refolding NMR⁷ and CIDNP pulse-labelling NMR⁸ experiments was further modified to improve its performance and durability. Details of the proteins and reagents and the injection device can be found in Supplementary Information and Supplementary Fig. 1.

Rapid refolding and NMR experiments. A volume of 30 μ l of 1.5–2.0 mM TC5b or ¹⁹F-Tyr-TC5b (a control, see Supplementary Information) in 6.0 M urea, 0.05 M sodium citrate buffer, pH 7.5, with 0.3 mM FMN was loaded as injectant. The experiment was performed at 5 °C to ensure essentially complete refolding to the native state⁵ (Supplementary Fig. 5). Polarization of the injectant was carried out using an Ar⁺ laser (Spectra Physics Beamlok 2080, multiline mode, 24 W output power) gated for a specified time period (usually 500 ms) using a mechanical shutter (NM Laser Products LS200). The unfolded protein solution was then pneumatically injected into 300 μ l of refolding buffer (0.05 M sodium citrate, pH 7.5) contained in the NMR sample tube, resulting in an 11-fold dilution of the denaturant. Complete mixing and folding is accomplished within 50 ms of injection⁷. (The first-order time constant for refolding of TC5b has been determined¹⁶ to be 4 μ s.) Each pulse-labelled spectrum was the difference between the sum of eight ‘light’ and ‘dark’ experiments, where ‘light’ spectra are those acquired after a laser flash, and ‘dark’ spectra are those acquired without laser illumination. Further details on the photo-CIDNP and conventional NMR experiments, along with data analysis and extraction of cross-polarized distances, are described in Supplementary Information.

Received 23 August 2006; accepted 2 March 2007.

Published online 11 April 2007.

1. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
2. Fersht, A. R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582 (2002).
3. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
4. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
5. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nature Struct. Biol.* **9**, 425–430 (2002).
6. Kaptein, R. in *Biological Magnetic Resonance* (eds Berliner, L. J. & Reuben, J.) 145–191 (Plenum, New York, 1982).
7. Mok, K. H. *et al.* A rapid sample-mixing technique for transient NMR and photo-CIDNP spectroscopy: Applications to real-time protein folding. *J. Am. Chem. Soc.* **125**, 12484–12492 (2003).

8. Mok, K. H., Nagashima, T., Day, I. J., Hore, P. J. & Dobson, C. M. Multiple subsets of side-chain packing in partially folded states of α -lactalbumins. *Proc. Natl Acad. Sci. USA* **102**, 8899–8904 (2005).
9. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622 (2004).
10. Klein-Seetharaman, J. *et al.* Long-range interactions within a nonnative protein. *Science* **295**, 1719–1722 (2002).
11. Shortle, D. & Ackerman, M. S. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487–489 (2001).
12. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**, 1341–1351 (2000).
13. Crowhurst, K. A. & Forman-Kay, J. D. Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain. *Biochemistry* **42**, 8687–8695 (2003).
14. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V. & Fersht, A. R. Solution structure of a protein denatured state and folding intermediate. *Nature* **437**, 1053–1056 (2005).
15. Balbach, J. *et al.* Detection of residue contacts in a protein folding intermediate. *Proc. Natl Acad. Sci. USA* **94**, 7182–7185 (1997).
16. Qiu, L., Pabit, S. A., Roitberg, A. E. & Hagen, S. J. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μ s. *J. Am. Chem. Soc.* **124**, 12952–12953 (2002).
17. Simmerling, C., Strockbine, B. & Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124**, 11258–11259 (2002).
18. Snow, C. D., Zagrovic, B. & Pande, V. S. The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.* **124**, 14548–14549 (2002).
19. Chowdhury, S., Lee, M. C., Xiong, G. & Duan, Y. *Ab initio* folding simulation of the Trp-cage mini-protein approaches NMR resolution. *J. Mol. Biol.* **327**, 711–717 (2003).
20. Pitera, J. W. & Swope, W. Understanding folding and design: Replica-exchange simulations of “Trp-cage” miniproteins. *Proc. Natl Acad. Sci. USA* **100**, 7587–7592 (2003).
21. Zhou, R. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl Acad. Sci. USA* **100**, 13280–13285 (2003).
22. Mok, K. H. & Hore, P. J. Photo-CIDNP NMR methods for studying protein folding. *Methods* **34**, 75–87 (2004).
23. Hore, P. J., Egmond, M. R., Edzes, H. T. & Kaptein, R. Cross-relaxation effects in the photo-CIDNP spectra of amino acids in proteins. *J. Magn. Reson.* **49**, 122–150 (1982).
24. Kohn, J. E. *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl Acad. Sci. USA* **101**, 12491–12496 (2004).
25. Neuweiler, H., Doose, S. & Sauer, M. A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proc. Natl Acad. Sci. USA* **102**, 16650–16655 (2005).
26. Ahmed, Z., Beta, I. A., Mikhonin, A. V. & Asher, S. A. UV-resonance Raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. *J. Am. Chem. Soc.* **127**, 10943–10950 (2005).
27. Bunagan, M. R., Yang, X., Saven, J. G. & Gai, F. Ultrafast folding of a computationally designed Trp-cage mutant: Trp²-cage. *J. Phys. Chem. B* **110**, 3759–3763 (2006).
28. Sánchez, I. E. & Kiefhaber, T. Hammond behavior versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **327**, 867–884 (2003).
29. Khan, F., Kuprov, I., Craggs, T. D., Hore, P. J. & Jackson, S. E. ¹⁹F NMR studies of the native and denatured states of green fluorescent protein. *J. Am. Chem. Soc.* **128**, 10729–10737 (2006).
30. Lyon, C. E., Jones, J. A., Redfield, C., Dobson, C. M. & Hore, P. J. Two-dimensional ¹⁵N-¹H photo-CIDNP as a surface probe of native and partially structured proteins. *J. Am. Chem. Soc.* **121**, 6505–6506 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Kuprov for the synthesis of 9-fluorenylmethoxycarbonyl-O-t-butyl-3-fluoro-L-Tyr and for NMR experiments on the ¹⁹F-Tyr-TC5b variant; T. Nagashima, C. J. V. Jones and H. Paisley for help with the conceptual design, building and testing of the rapid mixing injector; R. Gerber for assistance in acquiring the NMR spectra; A. L. Davis for discussions and for providing spectrometer time for the diffusion measurements; L. J. Smith, C. Redfield, A. E. Mark and D. A. C. Beck for discussions; and S. Min for assistance in figure preparation. K.H.M. also thanks M. Nilges, R. Wade and the EMBO Practical Course on Biomolecular Simulation. We are indebted to C. M. Dobson for continued encouragement in the application of photo-CIDNP to protein folding problems. This work was supported by the BBSRC (K.H.M., L.T.K., and P.J.H.), the Studienstiftung des deutschen Volkes (L.T.K.), the Deutsche Forschungsgemeinschaft (M.G.), and the US National Institutes of Health (N.H.A. and J.C.L.).

Author Contributions K.H.M. and I.J.D. built the *in situ* rapid mixing injector. K.H.M., M.G., I.J.D. and P.J.H. designed the experiments. K.H.M., L.T.K., M.G. and I.J.D. performed the experiments. J.C.L. and N.H.A. contributed the TC5b sample. M.G. developed the mathematical methods for obtaining NOE contact distances. K.H.M., L.T.K., M.G., I.J.D., N.H.A. and P.J.H. analysed the data. K.H.M., M.G., N.H.A. and P.J.H. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.J.H. (peter.hore@chem.ox.ac.uk) and K.H.M. (mok1@tcd.ie).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

When Bill Clinton was running for the US presidency in 1992, he focused on getting the top job with a campaign mantra coined by his strategist James Carville: "The economy, stupid!" Researchers looking to work in the biotechnology industry during the next year might consider adapting that refrain to "the pipeline, scientist!" in the light of consultants Ernst & Young's recent publication *Beyond Borders: Global Biotechnology Report 2007*.

Because the money available for start-ups and initial public offerings has decreased in recent years, the safest bet for finding a reasonably stable job in the sector is to target companies that have multiple products in different stages of development. Companies with at least one product approved for marketing are a safer bet than those without any. Of course, nothing is certain in a volatile sector — the number of mergers and acquisitions in the sector could well increase during the next 12 months, partly driven by drug companies that face expiring patents on some of their top performers.

But despite the uncertainties, biotech chief executives seem sanguine about job opportunities and growth. Global employment in the sector rose 30% to 190,500 between 2005 and 2006. Most of these workers (130,600) are based in the United States. Chief executives surveyed by Ernst & Young expect this growth to continue: 94% said they were likely or very likely to increase employment in the next two years.

But even with the projected hiring increase, there are still risks to weigh. Outsourcing in both sales and clinical trials is also expected to increase, and although mergers and acquisitions will usually bolster a company's finances, they tend to result in job cuts. By studying reports such as that from Ernst & Young, and then assessing indicators such as the state of the pipeline, and the number of patents and products already on the market, scientists can make smart choices about which jobs to go for. That could see them working in an industry that shows signs of growing towards profitability, a trend that should also mean greater job security.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Editor: Paul Smaglik

Assistant Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock

Naturejobs online production:

Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
e-mail: a.watanabe@natureasia.com

MOVERS

Klaus Stöhr, director of the influenza vaccine franchises, Novartis Vaccines and Diagnostics, Cambridge, Massachusetts



2006-07: Senior adviser, Influenza Pandemic Vaccine Development, World Health Organization (WHO), Geneva, Switzerland

2001-06: Coordinator, WHO Global Influenza Programme, Geneva, Switzerland

2003: Coordinator, SARS aetiology and diagnosis, WHO, Geneva, Switzerland

Many graduate students tackle complex research problems. Rarely, however, does a PhD dissertation help to craft a capable international response to a pandemic. Klaus Stöhr's did. Having previously secured his veterinary medicine degree at the University of Leipzig in what was then East Germany, Stöhr continued for a PhD focused on developing processes and procedures necessary to detect and identify the sources of emerging diseases in good time.

His thorough, proactive work soon landed him a position as director of a rabies vaccine development programme at the National Institute for Epidemiology and Infectious Disease Control in Animals. It would be the first of many projects capitalizing on his attention to detail and planning.

Leaders of the World Health Organization (WHO) took notice, and Stöhr accepted a job there in 1991. His initial work on rabies led to efforts focused on foodborne diseases, then antimicrobial resistance.

By 2000, he was charged with revitalizing the WHO influenza programme. Within two years, the programme reached a level of international recognition. David Heymann, the WHO's executive director for communicable diseases, credits Stöhr with not only developing the first WHO pandemic plan for influenza, but also building a cohesive network of cooperating laboratories around the world that could monitor for influenza and communicate rapidly. Indeed, Stöhr's existing network pinpointed the SARS corona virus in record time — four weeks after it first emerged in 2003. He went on to work with nations to develop global surveillance and laboratory methods, and he advised the pharmaceutical industry to help it transform the correct vaccine strains into vaccines. He has since been leading the WHO's response to avian influenza.

This year, after 15 years at the WHO, Stöhr has made what some may think is a surprising move. With offers on the table from the Harvard School of Public Health and several companies, he accepted a post as director of the influenza vaccine franchises at Novartis Vaccines and Diagnostics. He will focus on the business of vaccine development — the aspect of public health he found most challenging.

"In the public-health arena, there is not a good understanding of working with 'pharma', which is why I want to understand it better," says Stöhr.

Industry, says Heymann, is lucky to have Stöhr — whose excellent understanding of the basic science is only surpassed by his tireless efforts in public health. ■

Virginia Gewin

NETWORKS & SUPPORT

Gem of an idea for a unique machine

As summer approaches, universities across the globe will begin assembling teams of talented undergraduates from diverse disciplines to participate in the annual international genetically engineered machine (iGEM) competition. The contest tries to spark advancements in synthetic biology by encouraging research teams to design and construct new, unique genetic devices based on biological cells and circuits. The students, along with faculty members and graduate student mentors, are asked a simple question: "What would you like to build this summer?"

The iGEM team at the University of Texas, Austin, for example, engineered *Escherichia coli* to sense and respond to light in a novel way, enabling the construction of a bacteria-based photo-printing system (see *Nature* **438**, 441–442; 2005).

This year the Harvard iGEM team will consist of about a dozen students, mentored by faculty members, postdocs and graduate advisers. It is a rare opportunity for undergraduates to learn from and interact with a host of colleagues who are all interested in fostering the students' excitement for science and research.

Unlike traditional courses, the iGEM platform gives undergraduates much more freedom in the design,

direction and implementation of their projects, and helps build essential skills for team work. Students are encouraged to brainstorm and to propose ideas and experiments to pursue during the summer. These sessions not only build team spirit but also help the students to start thinking critically about how research is really done, from the initial idea to the experimental design.

Students work together throughout the summer. Then in November, teams from across the world come together to present and celebrate their work at the annual iGEM jamboree. Some projects inspire further development by future iGEM teams; others also play an important role by contributing directly to projects that postdocs and graduate students are working on.

The iGEM competition has been growing rapidly since it was launched in 2004, starting with 5 teams in 2004, 13 teams in 2005, and 37 teams in 2006. This year, it is estimated that 80 to 100 teams will participate. We encourage other universities to join us and use this unique platform for research and education. ■

Michael Strong is a Harvard Medical School postdoc and an iGEM teaching fellow. George Church is a Harvard Medical School professor of genetics and an iGEM faculty adviser.

POSTDOC JOURNAL

The balancing act

Recently, within the space of six weeks, I did a presentation at a department seminar, prepared a lab meeting, submitted an abstract for a scientific workshop, underwent the semi-annual review of my external funding, networked at a career-development event, conducted experiments for my project, drafted a manuscript for submission and, to top it all off, proofread both the primary data paper and the book chapter that my husband, also a postdoc, is writing.

Then, of course, there were personal matters to attend to, such as our imminently expiring apartment lease and potential homelessness.

At times like these, when my hours are long and my 'to do' lists even longer, I'm tempted to join the circus. After all, postdocs are expert jugglers: we constantly have multiple experiments, responsibilities and commitments up in the air. We are agile tightrope walkers: from start to finish, we have to balance our professional goals with our mentor's expectations, our family obligations and our personal priorities, sometimes without a safety net. We're capable contortionists: we demonstrate remarkable flexibility when put in a tight spot. Sometimes we're lion tamers, winning the cooperation and respect of those who can be notoriously challenging to work with. Most of all, we're dedicated performers. We know, always, that the show must go on. ■

Maria Thelma Ocampo-Hafalla is a research fellow at Cancer Research UK's London Research Institute.

The inside track from academia and industry

Dr Jekyll and Ms Hide

Where are the women in science? And what would attract them from other sectors?



Jan Bogg

Women are under-represented in science, engineering and technology (SET) posts in most economies. In Britain, there is one woman for every five men in SET occupations. Some constants are global — for example, even fewer women work in engineering than in other scientific disciplines. There are also some striking differences by sector, with women being concentrated in health and education — key employment areas for women generally. Women account for less than a tenth of the SET workforce in manufacturing and construction.

In higher education, about a fifth of all researchers, senior lectures and professorial staff are female. Interestingly, reports published in 2005 by the Engineering and Physical Sciences Research Council and other UK research bodies show similar levels of grant success rates for women and men.

Worryingly, women are not earning equal pay. Data from the UK Office for National Statistics in 2005 indicate that women still earn less than men across SET sectors, such as science, health, teaching, research, business and public service.

As in other areas of employment, managerial occupations remain strongly gender-segregated. Women are concentrated in traditional areas such as health and social services.

On a more positive note, the number of women in UK management roles has trebled during the past ten years, and women now account for 14% of company directors and 14% of managers in SET sectors.

This was the area in which my team did research with senior women in the bio- and health sciences during 2006. We interviewed 127 women in a variety of public and private-sector organizations.

They felt that, in order for others to progress, and in greater numbers, traditional models of management needed to be challenged throughout the organization. Equality and diversity training should be promoted.

How can employers achieve this? Employees require time investment, in terms of training and development to encourage and promote their growth. The training should address the employee's attitude and motivation. Also, there should be transparency in recruitment and promotion procedures. This must not just be in terms of policy — it must be known to be in practice throughout the organization.

The women we interviewed also named flexible working, work-life balance, networking and mentoring as important aspects for career progression.

Many employers now offer some form of flexible working, but consider your organization — is it more than a policy? Is it a reality? Do employees really feel that employers and colleagues endorse the policy? Employees may have outside commitments, so think about truly promoting flexible working wherever possible. Encourage some networking events within work hours, as some employees — men as well as women — have to care for children or elderly relatives. Working late may be difficult for them, making them feel they must let down either their colleagues or their dependants. Covering the cost of care by someone else could help.

When it comes to promotions, does your organization reward people for working long hours? Quality of work should be more important than hours spent in the workplace. Remember that allowing for a healthy work-life balance will provide benefits for the organization in terms of productivity, creativity and commitment. Do we really need a long-hours culture? If an

experiment has to be completed, a product launched or a paper finished, there may be a rationale for working long hours. But this should be the exception, not the norm. Flexibility should work in more than one direction.

Encourage a mentoring culture, train the mentors to provide a high-quality service and recognize that enforced mentoring does not always work, either because of personality clashes or reluctance by the

"Does your organization reward people for working long hours? Quality of work should be considered more important."

mentor or protégé. Importantly, provide the time for mentoring as part of an employee's existing workload, instead of as an add-on.

Mentors may be able to assist in guiding protégés into suitable networks. This would not only promote networking but also remove some of the mystique often associated with networks, especially specific ones such as those for women only.

Currently my work is focused on developing the lessons learned into frameworks for change in policy and practice. This includes new research evaluating mid-career women working in the bio- and health-sciences sectors. Issues to be addressed include barriers, drivers and progression to seniority. The challenge for employers is to listen to research evidence and to adapt to changing workforce needs. ■

Jan Bogg is director of the Breaking Barriers Project, University of Liverpool, UK. For more information on the project, to see executive summaries or to complete the online questionnaire, visit www.liverpool.ac.uk/breakingbarriers

"Mentoring should be part of an employee's workload, not an add-on."